



Yannis Tsividis

Operation and Modeling of
The MOS
Transistor

Second Edition

Selected List of Symbols

Certain symbols used only locally within a section, or whose meaning is clear from the context, are not included in this list.

| Symbol | Description | Section | Symbol | Description | Section |
|-----------------|---|----------|-------------|---|---------------|
| C'_b | Depletion region capacitance per unit area | 2.6 | I_D | Drain current | 1.6, 6.6, 7.2 |
| C_{bd} | Body-drain intrinsic capacitance | 8.3.2 | I_{DB} | Drain-to-body current | 6.6 |
| C_{bs} | Body-source intrinsic capacitance | 8.3.2 | I_{DS} | Drain-to-source current (channel current) | 4.3, 6.6 |
| C_{gb} | Gate-body intrinsic capacitance | 8.3.2 | I'_{DS} | Value of I_{DS} at the onset of saturation | 4.5.1 |
| C'_{gb} | Gate-body capacitance per unit area | 2.6 | I_{DSN} | Nonsaturation I_{DS} | 4.5.1 |
| C_{gd} | Gate-drain intrinsic capacitance | 8.3.2 | I_{DS1} | Component of I_{DS} due to drift | 4.3 |
| C_{gs} | Gate-source intrinsic capacitance | 8.3.2 | I_{DS2} | Component of I_{DS} due to diffusion | 4.3 |
| C'_i | Inversion layer capacitance per unit area | 2.6 | I_G | Gate current | 4.1, 7.2 |
| C'_{it} | Interface traps capacitance per unit area | 2.6 | I_M | Value of I_{DS} at upper limit of weak inversion | 8.3.2 |
| C_j | pn junction capacitance | 1.5 | I'_M | Value of I_{DS} at upper limit of weak inversion, normalized to (W/L) | 4.6 |
| C'_j | pn junction capacitance per unit area | 1.5 | I_S | Source current | 7.2 |
| C_m | Difference between C_{dg} and C_{gd} | 9.2.1 | I_T | Transport current | 7.2 |
| C_{mb} | Difference between C_{db} and C_{bd} | 9.2.1 | I_Z | Characteristic current in moderate inversion | 8.2.7 |
| C_{mx} | Difference between C_{bg} and C_{gb} | 9.2.1 | i_{DA} | Charging component of drain current | 7.3 |
| C_{ox} | Total oxide intrinsic capacitance | 8.3.2 | i_{SA} | Charging component of source current | 7.3 |
| C'_{ox} | Oxide capacitance per unit area | 2.2 | k | Boltzmann's constant | 1.2 |
| d_B | Depletion region depth | 2.5 | L | Effective channel length | 1.6 |
| \mathcal{E} | Electric field | | l_p | Length of pinchoff region | 6.2 |
| \mathcal{E}_c | "Critical" field in velocity saturation formulation | 6.5 | N_A | Acceptor concentration | 1.2 |
| \mathcal{E}_x | Longitudinal electric field | 6.5 | N_D | Donor concentration | 1.2 |
| \mathcal{E}_y | Transverse electric field | 4.10 | n | 1. Free electron concentration | 1.2 |
| g_{bd} | Body-drain small-signal conductance | 8.2.3 | | 2. The quantity $(d\psi_{sa}/dV_{GB})^{-1}$ | 2.5 |
| g_m | Gate small-signal transconductance | 8.2.2 | n_i | Intrinsic carrier concentration | 1.2 |
| g_{mb} | Body small-signal transconductance | 8.2.2 | p | Hole concentration | 1.2 |
| g_o | Output small-signal conductance | 8.2.2 | Q_B | Depletion region charge | 7.2 |
| g_{sd} | Source-drain small-signal conductance | 8.2.2 | Q'_B | Depletion region charge per unit area | 2.5 |
| g_{ss} | Source small-signal conductance | 8.2.2 | \hat{Q}_B | Effective depletion region charge | 6.3 |
| I_B | Body current | 4.1, 7.2 | Q_I | Inversion layer charge | 2.5, 7.2 |
| | | | Q'_I | Inversion layer charge per unit area | 2.5 |
| | | | Q'_{IL} | Value of Q'_I at drain end of channel | 4.3 |
| | | | Q'_{I0} | Value of Q'_I at source end of channel | 4.3 |
| | | | Q_G | Gate charge | 2.5, 7.2 |
| | | | Q'_G | Gate charge per unit area | 2.5 |
| | | | Q'_o | Effective interface charge | 2.2 |
| | | | q | Magnitude of electronic charge | 1.2 |
| | | | q_D | Drain-associated inversion layer charge | 7.3 |

| Symbol | Description | Section | Symbol | Description | Section |
|-------------|---|--------------|------------------|---|---------------|
| q_s | Source-associated inversion layer charge | 7.3 | y | Distance in direction perpendicular to the surface | 2.5 |
| T | Absolute temperature | 1.2 | y_m | Gate transadmittance | 9.3 |
| t_{ox} | Oxide thickness | 2.2 | y_{mb} | Body transadmittance | 9.3 |
| V_A | Characteristic voltage in first-order channel length modulation formula | 6.2 | W | Effective channel width | 1.6 |
| V_B | Body voltage | 7.2 | α | Coefficient of first-order term in expansion for $-Q'_B/C'_{ox}$ | 4.3.2, 4.5.3 |
| V_{CB} | Channel-body voltage | 3.2 | α_1 | Value of α for expansion around the source potential | 4.3.2, 4.5.3 |
| V_D | Drain voltage | 7.2 | γ | Body effect coefficient | 2.5, 3.3 |
| V_{DS} | Drain-source voltage | 1.6, 4.1 | Δ | Symbol denoting a change in the quantity following it | |
| V'_{DS} | Value of V_{DS} at onset of saturation | 4.5.3 | $\Delta\phi$ | Difference between the actual strong-inversion surface potential and its classical value of $2\phi_F$ | 2.5 |
| V_{FB} | Flat band voltage | 2.2 | ϵ_o | Permittivity of free space | 1.2 |
| V_G | Gate voltage | 7.2 | ϵ_{ox} | Permittivity of SiO_2 | 2.2 |
| V_{GS} | Gate-source voltage | 1.6, 4.1 | ϵ_s | Permittivity of silicon | 2.2 |
| V_H | Value of V_{GC} , or of V_{GS} , at onset of strong inversion | 3.4, 4.4 | η | Degree of nonsaturation | 4.5.3 |
| V_{HB} | Value of V_{GB} at onset of strong inversion | 3.4 | μ, μ_{eff} | Effective surface mobility | 4.3, 4.10 |
| V_{HO} | Value of V_{HB} for two-terminal structure | 2.5 | μ_B | Bulk mobility | 1.3 |
| V_L | Value of V_{GC} , or of V_{GS} , at onset of weak inversion | 3.4, 4.4 | τ | Transit time | 1.2, 7.5 |
| V_{LB} | Value of V_{GB} at onset of weak inversion | 3.4 | ϕ_{bi} | Built-in potential pn junction | 1.5 |
| V_{LO} | Value of V_{LB} for two-terminal structure | 2.5 | ϕ_F | Fermi potential | 1.4 |
| V_M | Value of V_{GC} , or of V_{GS} , at onset of moderate inversion | 3.4, 4.4 | ϕ_{MS} | Contact potential of body material to gate material ("work function difference" potential) | 2.2 |
| V_{MB} | Value of V_{GB} at onset of moderate inversion | 3.4 | ϕ_i | Thermal voltage, kT/q | 1.2 |
| V_{MO} | Value of V_{MB} for two-terminal structure | 2.5 | ϕ_Z | Moderate inversion region width in terms of surface potential | 3.4 |
| V_P | Pinchoff voltage | 3.5, 4.5 | ϕ_{Z0} | Moderate inversion region width in terms of surface potential for two-terminal MOS structure | 2.5 |
| V_Q | Value of channel-body voltage at the boundary between strong and moderate inversion, for a given V_{GB} | 3.5, 4.4 | ϕ_0 | Surface potential of two-terminal MOS structure in strong inversion | 2.5 |
| V_S | Source voltage | 7.2 | ψ_{ox} | Oxide potential | 2.2 |
| V_{SB} | Source-body voltage | 1.6, 4.1 | ψ_s | Surface potential | 2.2 |
| V_T | Extrapolated threshold voltage in terms of V_{GC} or V_{GS} | 3.4.2, 4.5.3 | ψ_{sa} | Surface potential in the absence of inversion layer, for a given V_{GB} | 2.5, 3.2, 4.6 |
| \hat{V}_T | Effective V_T | 6.3.2 | ψ_{sL} | Value of surface potential at drain end of channel | 4.3 |
| V_{TB} | Extrapolated threshold voltage in terms of V_{GB} | 3.4.2 | ψ_{s0} | Value of surface potential at source end of channel | 4.3 |
| V_{TO} | Value of V_{TB} for two-terminal structure | 2.5.2 | ω_o | Characteristic angular frequency | 8.3.2 |
| V_W | Value of channel-body voltage at the boundary between weak inversion and depletion, for a given V_{GB} | 3.5, 4.4 | ω_{Ti} | Intrinsic transition angular frequency | 8.3.2 |
| v_d | Drift velocity | 1.3, 6.5 | | | |
| x | Distance along the channel | 4.1 | | | |

Values for some useful quantities

| | |
|---|--|
| Magnitude of electronic charge, q | $1.602 \times 10^{-19} \text{ C}$ |
| Thermal voltage, $\phi_t = kT/q$, at 300 K | 0.0259 V |
| Permittivity of silicon, ϵ_s | $1.04 \times 10^{-12} \text{ F/cm}$ |
| Permittivity of silicon dioxide, ϵ_{ox} | $3.45 \times 10^{-13} \text{ F/cm}$ |
| $\sqrt{2q\epsilon_s}$ | $5.79 \times 10^{-16} \text{ F} \cdot \text{V}^{1/2} \cdot \text{cm}^{-1/2}$ |

Operation and Modeling of

The MOS Transistor

Operation and Modeling of

The MOS Transistor

Second Edition

Yannis Tsividis

Columbia University

New York Oxford
OXFORD UNIVERSITY PRESS

This book provides a unified treatment of the many phenomena encountered in the operation of modern MOS transistors, and shows how such phenomena can be modeled analytically. The book is mainly written for use in a senior or first-year graduate course. It is felt that electrical engineering students have much to gain from a course devoted to the subject. The MOS transistor is the dominant VLSI device. A course devoted to it is, of course, invaluable to those planning a career in device physics and modeling. For such people, the standard courses on semiconductor devices usually cover too many different devices to do justice to any one of them, and do not present the intricacies and tradeoffs involved in a detailed modeling effort. The value of a course devoted to the MOS transistor is also extremely high for those who want to use the device to design state-of-the-art circuits. Integrated circuit designers have the opportunity to suit devices to circuit needs, and they can do this most intelligently if they really understand the workings of the devices. One can, of course, design systems by using predesigned circuit building blocks as black boxes, if truly high performance is not important. But when state-of-the-art performance is a must, one has to consider device details. In addition, a deep knowledge of device operation and modeling is needed for understanding the computer simulator models a designer is working with, and for identifying their limitations. Many circuit designers in the industry spend endless hours trying to interpret strange circuit simulation results, not realizing that these are largely due to modeling inadequacies. Without adequate device understanding, valuable time and effort is bound to be wasted on overdesign, brute-force approaches, and design iterations. This author believes that no IC designer's education is complete without detailed exposure to MOS transistor operation and modeling.

In the dozen years that have passed since the publication of the first edition of this book there have been significant advances in the understanding and modeling of the MOS transistor. In addition, the requirements for modeling this device on the part of the circuit design community are now much more demanding. For example, the push for low-voltage and micropower operation has made necessary careful modeling of the device below strong inversion, and the push for ever smaller dimensions has

revealed phenomena previously ignored. In addition, the advent of "mixed-signal" circuits, which combine analog and digital functions on the **same chip**, makes it **necessary** to use models that are good enough for **analog** work. These **developments** have pointed to the need for a major revision of this book. However, the basic **philosophy** of the first edition has been retained. Several aspects of this philosophy are summarized below.

- The **book** starts with basic concepts. Readers **should** be able to follow even if they had no prior exposure to the device. The discussions of these concepts are often from a **perspective** different from the one usually taken, thus **making** them interesting reading even to those with prior **exposure**.
- Every **effort has** been made to give the subject a careful treatment. The reader may at times get the feeling that the author is "splitting hairs." The author **would rather** be **accused** of this **than fudge**. The MOS transistor **is a** device so complex that, once one has **decided** to fudge, things grow out of hand very quickly, and one ends up with a hodgepodge of careless derivations, conflicting models, and a lot of **patchwork**. This **has** been avoided at all costs. Also, the reason for hair-splitting at **some points** can only be appreciated further along in the **book**, where those "too fine" **details can** be seen to make a lot of difference. The use of one name for several quantities, common in some of the literature, is carefully avoided. For **example**, at least four distinct quantities encountered in MOS transistor work are **described by using** the single name "threshold voltage." **Although** the reader is **amply** warned of this practice, the practice itself is **avoided** in **this book**.
- The emphasis is on principles. **At** the same time, to illustrate these principles, relevant models are **extensively derived** and **discussed**. Thus, **physics** and modeling are discussed in **parallel** throughout the book.
- Analytical results are derived **in a** logical manner after carefully stating the **assumptions made**. Empirical modeling is avoided as much as **possible**. **However**, **there are phenomena** for which the only analytical results available are empirical or semiempirical. Such results are presented for completeness after pointing out the necessary hand-waving behind them.
This book is **not a** survey. In fact, a conscious effort has been made to avoid **making** it **one**. A well-connected set of topics has been chosen, and **most** of these are **discussed** in significant detail. Nevertheless, for completeness certain other topics **are** mentioned, albeit briefly. In such cases, some representative results are shown without proof, so that the reader can know what to expect if he or she consults the references provided.
- A **great deal of** emphasis is placed on providing intuition for the various **phenomena discussed**. It is rather hopeless to attempt working with a device as complex as the MOS transistor relying only on **analytical** relations. The emphasis on **intuition** has made lengthy discussions necessary.
- The **pace is** unhurried. The author **believes** that this actually makes it possible to study the material faster. Thus, whereas the treatment of a **given topic may be long** in terms of number of pages, it actually should take **less time** to comprehend it, **because** of both the detailed **derivations and the** intuitive discussions. **At** times, the

reader **may** get the **feeling** of *déjà vu*, since some points are repeated more than once **to make** sure they are not missed, especially if the reading of the topics **is** done out of sequence. In **general**, the **book is written in the style** in which the author would like to have any new subject presented to him. He would very much like to see the new subject "beaten to death:" presented with several points of view to increase perspective and with a significant amount of repetition. The author has been in the past **grateful** for treatments **of** this type and never felt offended by this **style**. If the reader happens to be "faster" than **the author** in this **respect**, he or **she** can easily skip some of the discussions.

Almost all chapters in this book, and **almost all sections** within **each chapter**, have been **extensively revised**. Many sections **have** been rewritten, and **new ones have been added**. **One** chapter is entirely new. A list of chapters follows, along with an **explanation** of their features and of what is **new** in them in this **edition**.

Chapter 1: Semiconductors, Junctions, and MOSFET Overview

All preliminary material necessary for the understanding of MOS structures is given here. **This material is** important to the newcomer, **but part of it should also make interesting reading** for **those** with some previous exposure to basics. This includes the **material on contact potentials**, which is used to advantage in the following chapter. The chapter concludes **with an overview of the** MOS transistor. This section is **new** to this **edition**. It **provides a framework for the rest of the book, and makes it clear why** particular details of the **two- and three-terminal MOS structures are studied in the following two chapters**.

Chapter 2: The Two-Terminal MOS Structure

Here the reader **will** find a treatment of the MOS structure with gate and substrate **terminals** only. Concepts not directly related **to** the presence of the source and **drain** in the **MOS** transistor are treated here. The regions of weak, moderate, and strong inversion **are** all introduced in this chapter. **Potentials** are used throughout rather than energy **bands**. **This is not only** common **in current literature** but also **helps** provide rigorous straightforward derivations. Consider, for example, the well-known term ϕ_{MS} **appearing in the expression** for the **flat-band** voltage. In energy band **treatments** it is **often not** clear where in the MOS structure this potential actually **resides**. **In this book**, it is made evident that ϕ_{MS} is nothing but a contact potential, and the places where it resides are made obvious. Also, its presence in the flat-band voltage **expression is rigorously** justified through Kirchhoff's voltage law. The material on **weak and moderate inversion has been** improved and expanded in this **edition**.

Chapter 3: The Three-Terminal MOS Structure

Here one more terminal is added to the structure of Chap. 2, to connect the **inversion** layer to the external world. MOS transistor concepts that are not directly related to

current flow are presented in this chapter. This includes the important "substrate effect," which is amply treated. The section on limits of regions of inversion has been streamlined and shortened. On the other hand, an entirely new section, "A ' V_{CB} Control' Point of View," has been added, to lay the foundation for the discussion of certain recent models covered in Chap. 4.

Chapter 4: The Four-Terminal MOS Transistor

The four-terminal MOS transistor is obtained in this chapter by adding one terminal to the structure of Chap. 3. This device is now very easy to understand, on the basis of the concepts already presented for the two- and three-terminal structures. This is the central chapter in the book. Several models are presented in detail. The first of them is the complete charge sheet model, including drift and diffusion currents, valid in all regions of operation. Thanks to a simplified derivation, this material is brief but thorough. This is followed by new material on simplified charge sheet models, including both symmetric and source-referenced versions.

The above models form the basis for deriving several popular strong- and weak-inversion models, which are covered in detail. A considerable part of this material is new to this edition, and has been included to reflect recent trends. Some of this material can be skipped without loss of continuity, and this is indicated at the appropriate points. The various models are extensively related and/or compared to each other, and the way they can all be derived from one master model (the complete charge sheet model) is pointed out. A new section on interpolation models has been added. Sections on effective mobility (expanded), temperature effects, etc., are also included. The tradeoffs between accuracy and simplicity are pointed out throughout the chapter.

Chapter 5: MOS Transistors with Ion-Implanted Channels

This chapter was Chap. 6 in the first edition. It now precedes the chapter on small dimension effects. This change was made because all modern devices, small or large, have ion-implanted channels; also, having been exposed to this material, the reader can understand better certain small-dimension effects discussed in the following chapter. Nevertheless, Chaps. 5 and 6 have been revised in such a way that they can be covered in either order, in order to accommodate the need of instructors who prefer the original order.

This chapter, arguably the most tersely written one in the first edition, has been extensively revised. It has now been written so that specific sections correspond more closely to actual devices (enhancement nMOS, depletion nMOS, surface- or buried-channel pMOS), and a much smoother development is given. The revision also makes it possible, if desired, for an instructor to cover only the parts which discuss the effects of ion implantation on threshold voltage, and to skip the detailed development of other aspects of I-V characteristics. In certain settings, this may be necessary because of time limitations.

Chapter 6: Small-Dimension Effects

This **chapter** has been **revised and expanded** by an **expert** in device miniaturization, Prof. D. A. **Antoniadis** of MIT. The new title **reflects** the fact that the effects of **miniaturization** in all three dimensions (including very thin oxide effects) are discussed. Among the new topics in this edition are reverse short-channel and narrow-channel effects and hot carrier effects. Although space does not allow for a detailed exposition of all small-dimension effects, the reader is made aware of their **existence**, **and a qualitative** discussion is given. This includes such effects as poly gate depletion, nonzero inversion layer thickness, quantum mechanical threshold increase, and insulator tunneling.

Chapter 7. The MOS Transistor in Dynamic Operation — Large-Signal Modeling

This chapter is largely devoted to charge modeling. The concept of quasi-static operation is carefully introduced, and general techniques for charge evaluation are presented, **illustrated by charge** computations for one representative model. Non-quasi-static analysis is then introduced. Since this chapter was considered by instructors and reviewers to be one **of** the most successful ones in the first **edition**, its basic structure was retained. Some material was added on **general** charge modeling independent of inversion regions, and on transient response in non-quasi-static operation.

Chapter 8: Small-Signal Modeling for Low and Medium Frequencies

The principles behind small-signal modeling are presented. The discussion is limited **to** quasi-static **behavior**. A useful small-signal model is developed for operation at low and medium frequencies. Major changes in this chapter include a detailed discussion of the effects of substrate current, notably on output conductance, a discussion of single-piece expressions for small-signal parameters valid in all regions of inversion, an expanded discussion of small-dimension effects, and an expanded discussion of extrinsic capacitance modeling. Noise is then discussed, including an expanded description **of flicker** noise and of the effects of small device dimensions on **noise**.

Chapter 9: High-Frequency Small-Signal Models

In this chapter, two kinds of small-signal models are developed. First, complete **quasi-static** models are introduced, which differ from the models of **Chap. 8** in that they include transcapacitors. The **nature** of these somewhat controversial elements is carefully discussed. Techniques are given for the rigorous development of equivalent-circuit topologies from a complete quasi-static description. Then, non-quasi-static models are introduced through a careful development of the transistor's "transmission line" equations, and a useful **y-parameter** model is derived for high-frequency applications. It is shown that each **level** of modeling reduces to the next lower one if the

frequency is sufficiently reduced. In this edition, the section on **non-quasi-static** modeling **has** been revised to include a more extensive explanation for the presence of inductance in intrinsic small-signal models. A new section on high-frequency noise, emphasizing induced gate noise, has been added. Another **new** section deals **with** considerations for radio-frequency (RF) modeling, including the effects of gate resistance and discussing the two common figures of merit for high-frequency performance, the transition frequency **and** the maximum **frequency** of oscillation.

Chapter 10: MOSFET Modeling for Circuit Simulation

This chapter is entirely **new** in the second edition. and replaces the one on fabrication. Having been exposed **to** the many phenomena in the MOS transistor and **to the** modeling **of** such phenomena, the reader will find in this **chapter** an **exposition of the many** issues **and** considerations involved **in** putting all these together to **make an extensive** model suitable for circuit simulation. Discussed here are the various types of **models**, the ways that models of particular phenomena are combined, parameter extraction. **desirable** properties for simulator models, common pitfalls in modeling, and many benchmark tests for models, which **have** recently been included in an IEEE standard. This chapter, it is hoped, will provide a starting point for readers who **intend to work** in modeling, and will **save them time and effort by clearly** warning of common errors. **it is** also meant **to** provide a background for circuit designers, **allowing them to understand the** limitations of **the** models **they** are **using**, **and** to better communicate their needs to modeling **experts**.

The book concludes with 13 appendixes containing an introduction to energy band concepts, the basic laws of electrostatics as well as several general but complicated results which, it was **felt**, would distract **if put in the main text**, **For the same reason**, some material in the main text **was** put in fine **print or in** footnotes or, as already mentioned, was described in the statements of some homework problems, **Such material** includes certain fine details. alternative points of view, etc. To avoid **distraction**, **the** reader may prefer to skip fine-print footnotes and **appendixes** during a first reading; the main text is **self-contained**. This material can always be consulted at a later time, because its connection with specific **points** in the **text** is obvious.

References to the technical literature were **extensively** updated **and** expanded. In **most cases**, a reference was selected for inclusion because **it** is technically important, or is widely mentioned in the literature, or has historical significance, or is **part of controversy** that has not yet been resolved.

A change from the first edition has to do with units. The consistent system of units used in that edition has been abandoned, **as it failed to work well** in most settings. Thus, in the present edition common units have been adopted [e.g., \AA for oxide thickness and $\text{cm}^2/(\text{V}\cdot\text{s})$ for mobility]. In some instances, the value of a **quantity** is given **in such** units, with the value in different units given after that in parentheses; **for example**, a field intensity is given as “ $3 \times 10^4 \text{ V/cm}$ (or $3 \text{ V}/\mu\text{m}$),” **since the latter** form relates directly **to** typical values of voltages and channel lengths.

The subject of this book is definitely among those that are best digested by doing. The homework problems should help to give a feeling for the kind of work involved in modeling. Most of the homework problems fall in one of the following categories:

1. They sketch additional modeling ideas not in the main text and encourage the reader to try them out.
2. They compare several models introduced in the text.
3. They ask for computations and plots to help provide a quantitative feeling and investigate various properties.
4. They ask for detailed derivations which were sketched in the main text, but which were not shown in detail in order to avoid distraction from the main points.

Students can be encouraged to write subroutines for the calculation of various quantities as they go along and to save them for later use. In this way, they will gradually build a library of useful subroutines that can be helpful not only in new homework assignments but also in a final project if one is assigned. It has been the author's experience that project work is invaluable, and the more extensive the project, the greater the benefits. The project can take the form of the implementation of models on the computer. Here it is not enough to just copy a model from the book into a computer program. One must worry about choosing the right models, appropriately combining them to form general models, ensuring continuity of calculated quantities with respect to all given parameters, etc. Some examples of projects follow.

- Write a computer program to evaluate the drain current of a device on a uniform substrate, including short- and narrow-channel effects. The current should be continuous with respect to any input parameter (voltage, geometric dimensions, etc.), and so should be the derivatives of the current with respect to each terminal voltage.
- Develop a computationally efficient technique for the evaluation of drain current in a long-channel device, valid in weak, moderate, and strong inversion. This will necessitate the development of efficient numerical techniques, because the general charge sheet approach, if unmodified, will lead to complex computations. Again, continuity of the current and its derivatives with respect to all parameters should be ensured.
- Develop a program for modeling ion-implanted devices in strong inversion, again ensuring continuity.
- Develop a program for the modeling of low- and medium-frequency small-signal parameters in strong inversion, paying special attention to the small-signal output conductance in the saturation region. All small-signal parameters should be continuous with respect to all input parameters.
- Develop a program for the evaluation of all charges and small-signal capacitances. All these quantities should be continuous with respect to any input parameter.
- Develop a program for the extraction of parameter values to be used with a given model. The input to this program is assumed to consist of measured quantities.

In addition, certain long homework problems can easily be expanded into projects. In all cases, the value of the project, and the challenge in it, can be enhanced if the results are compared to measurements, obtained either in the lab or from the technical literature (the references provided should be very helpful in the latter case). Depending on the magnitude of effort foreseen, students can work separately or in teams.

The first edition of this book has been used for senior or graduate courses at many universities and in industrial short courses. For a one-semester course, a large number of possibilities exist in regard to the topics selected for coverage. Thus, for example, a course emphasizing general principles would cover Chaps. 1 through 3, Chap. 4 including a careful coverage of the complete charge sheet model, Chaps. 6 to 9, and selected topics from other chapters, depending on interest and time available. A course emphasizing practical models for digital circuit design may cover Chaps. 1 through 4, deemphasizing the general charge sheet model and the weak-inversion region, parts of Chap. 5, Chaps. 6 and 7, and Sec. 8.4. For parts that are deemphasized, a quick qualitative coverage is possible, based mostly on the figures. The author would be happy to consider individual teaching needs and suggest specific course outlines to instructors who contact him.

The author would like to acknowledge the contributions of many individuals to the shaping of this edition. He owes many thanks to his friend and collaborator Dimitri Antoniadis, for writing the chapter on small-dimension effects and for useful comments on other parts of the book. He is indebted to Colin McAndrew for his extremely extensive comments, suggestions, and his support of the first edition of the book, and to Narendra Rao for his detailed comments. Many thanks are due to Ana Isabela Cunha, Christian Enz, Mehmet Ozturk, Renan Türkman, and J. P. Sun for their reviews of parts of the manuscript for this edition. For their comments during the preparation of either this or the first edition, the author would like to also thank Pratul Ajmera, Narain Arora, Mehran Bagheri, Alexis Birbas, John Brews, Howard Card, Paul Diament, V. Gopinathan, Renuka Jindal, Ping-Keung Ko, Andy Lish, Gerson Machado, Ranjit Mand, Guido Masetti, Ed Nicolian, Shanthi Pavan, Charles Sordini, Ken Shepard, Ken Suyama, Don Ward, and Ed Yang. The author also thanks his editors, Lynn Cox and Brad Kosirog, for their enthusiasm and support, and Elsa Sanchez for text processing.

Yannis Tsividis

| | | |
|----------------------|---|---------------|
| Chapter 1 | Semiconductors, Junctions, and MOSFET Overview | 1 |
| 1.1 | Introduction | 1 |
| 1.2 | Semiconductors | 1 |
| 1.3 | Conduction | 8 |
| 1.3.1 | Transit Time | 8 |
| 1.3.2 | Drift | 9 |
| 1.3.3 | Diffusion | 15 |
| 1.4 | Contact Potentials | 18 |
| 1.5 | The <i>pn</i> Junction | 26 |
| 1.6 | Overview of the MOS Transistor | 34 |
| 1.6.1 | Basic Structure | 34 |
| 1.6.2 | A Qualitative Description of MOS Transistor Operation | 39 |
| 1.6.3 | A Fluid Dynamical Analog | 41 |
| 1.6.4 | MOS Transistor Characteristics | 44 |
| 1.7 | A Brief Overview of This Book | 46 |
| | References | 48 |
| | Problems | 49 |
| Chapter 2 | The Two-Terminal MOS Structure | 50 |
| 2.1 | Introduction | 50 |
| 2.2 | The Flat-Band Voltage | 51 |
| 2.3 | Potential Balance and Charge Balance | 56 |
| 2.4 | Effect of Gate-Substrate Voltage on Surface Condition | 58 |
| 2.4.1 | Flat-Band Condition | 59 |
| 2.4.2 | Accumulation | 59 |

| | |
|---|------------|
| 2.4.3 Depletion and Inversion | 59 |
| 2.4.4 General Analysis | 62 |
| 2.5 Inversion | 64 |
| 2.5.1 General Relations and Regions of Inversion | 64 |
| 2.5.2 Strong Inversion | 71 |
| 2.5.3 Weak Inversion | 74 |
| 2.5.4 Moderate Inversion | 78 |
| 2.6 Small-Signal Capacitance | 79 |
| 2.7 Summary of Properties of the Regions of Inversion | 87 |
| References | 87 |
| Problems | 88 |
| Chapter 3 The Three-Terminal MOS Structure | 91 |
| 3.1 Introduction | 91 |
| 3.2 Contacting the Inversion Layer | 92 |
| 3.3 The Body Effect | 101 |
| 3.4 Regions of Inversion | 103 |
| 3.4.1 Approximate Limits | 103 |
| 3.4.2 Strong Inversion | 106 |
| 3.4.3 Weak Inversion | 109 |
| 3.4.4 Moderate Inversion | 113 |
| 3.5 A " V_{CB} Control" Point of View | 113 |
| 3.5.1 Fundamentals | 113 |
| 3.5.2 "Pinchoff" Voltage | 117 |
| 3.5.3 Expressions in Terms of the "Pinchoff" Voltage | 120 |
| References | 122 |
| Problems | 123 |
| Chapter 4 The Four-Terminal MOS Transistor | 125 |
| 4.1 Introduction | 125 |
| 4.2 Transistor Regions of Operation | 128 |
| 4.3 General Charge Sheet Models | 131 |
| 4.3.1 Complete Charge Sheet Model | 131 |
| 4.3.2 Simplified Charge Sheet Models | 140 |
| 4.3.3 Model Based on Quasi-Fermi Potentials | 146 |
| 4.4 Regions of Inversion in Terms of Terminal Voltages | 147 |
| 4.5 Strong Inversion | 150 |
| 4.5.1 Complete Symmetric Strong-Inversion Model | 150 |
| 4.5.2 Simplified Symmetric Strong-Inversion Model | 156 |
| 4.5.3 Simplified, Source-Referenced, Strong-Inversion Model | 158 |
| 4.5.4 Model Origin Summary | 169 |
| 4.6 Weak Inversion | 170 |
| 4.7 Moderate Inversion | 175 |

| | | |
|------------------|--|------------|
| 4.8 | Interpolation Models | 176 |
| 4.9 | Source-Referenced vs. Body-Referenced Modeling | 179 |
| 4.10 | Effective Mobility | 181 |
| 4.11 | Temperature Effects | 189 |
| 4.12 | Breakdown | 191 |
| 4.13 | The <i>p</i> -Channel MOS Transistor | 192 |
| 4.14 | Enhancement-Mode and Depletion-Mode Transistors | 194 |
| 4.15 | Model Parameter Values, Model Accuracy, and Model Comparison | 195 |
| | References | 197 |
| | Problems | 203 |
| Chapter 5 | MOS Transistors with Ion-Implanted Channels | 207 |
| 5.1 | Introduction | 207 |
| 5.2 | Enhancement nMOS Transistors | 208 |
| 5.2.1 | Preliminaries | 208 |
| 5.2.2 | Charges and Threshold Voltages | 212 |
| 5.2.3 | Drain-to-Source Current Model for Strong Inversion | 219 |
| 5.2.4 | Simplified Model for Strong Inversion | 222 |
| 5.2.5 | Weak Inversion | 225 |
| 5.3 | Depletion nMOS Transistors | 225 |
| 5.3.1 | The Need for an <i>n</i> -Type Implant | 225 |
| 5.3.2 | Charges and Threshold Voltage | 227 |
| 5.3.3 | Transistor Operation | 234 |
| 5.4 | Enhancement pMOS Transistors | 240 |
| 5.4.1 | Surface-Channel Enhancement-Mode pMOS | 240 |
| 5.4.2 | Buried-Channel Enhancement-Mode pMOS | 241 |
| | References | 241 |
| | Problems | 245 |
| Chapter 6 | Small-Dimension Effects | 248 |
| | by D. Antoniadis, Massachusetts Institute of Technology | |
| 6.1 | Introduction | 248 |
| 6.2 | Channel Length Modulation | 249 |
| 6.3 | Barrier Lowering, Two-Dimensional Charge Sharing, and Threshold Voltage | 257 |
| 6.3.1 | Introduction | 257 |
| 6.3.2 | Short-Channel Devices | 257 |
| 6.3.3 | Narrow-Channel Devices | 270 |
| 6.3.4 | Summary and Comments | 277 |
| 6.4 | Punchthrough | 277 |
| 6.5 | Carrier Velocity Saturation | 280 |
| 6.6 | Hot Carrier Effects—Substrate Current, Gate Current, and Breakdown | 286 |
| 6.7 | Scaling | 290 |

| | | |
|------------------|---|------------|
| 6.8 | Effect of Surface and Drain Series Resistances | 296 |
| 6.9 | Effects due to Thin Oxides and High Doping | 298 |
| | References | 301 |
| | Problems | 309 |
| Chapter 7 | The MOS Transistor in Dynamic Operation— Large-Signal Modeling | 312 |
| 7.1 | Introduction | 312 |
| 7.2 | Quasi-Static Operation | 313 |
| 7.3 | Terminal Currents in Quasi-Static Operation | 317 |
| 7.4 | Evaluation of Charges in Quasi-Static Operation | 325 |
| 7.4.1 | Introduction | 325 |
| 7.4.2 | Strong Inversion | 325 |
| 7.4.3 | Moderate Inversion | 331 |
| 7.4.4 | Weak Inversion | 332 |
| 7.4.5 | General Charge Sheet Model | 333 |
| 7.4.6 | Depletion | 335 |
| 7.4.7 | Accumulation | 336 |
| 7.4.8 | Plots of Charges versus V_{GS} | 336 |
| 7.4.9 | Use of Charges in Evaluating Terminal Currents | 337 |
| 7.5 | Transit Time under DC Conditions | 339 |
| 7.6 | Limitations of the Quasi-Static Model | 340 |
| 7.7 | Non-Quasi-Static Modeling | 347 |
| 7.7.1 | Introduction | 347 |
| 7.7.2 | The Continuity Equation | 347 |
| 7.7.3 | Non-Quasi-Static Analysis | 348 |
| | References | 354 |
| | Problems | 358 |
| Chapter 8 | Small-Signal Modeling for Low and Medium Frequencies | 359 |
| 8.1 | Introduction | 359 |
| 8.2 | A Low-Frequency Small-Signal Model for the Intrinsic Part | 360 |
| 8.2.1 | A Two-Path View | 360 |
| 8.2.2 | A Small-Signal Model for the Channel Path | 361 |
| 8.2.3 | A Small-Signal Model for the Drain-to-Substrate Path | 365 |
| 8.2.4 | Strong Inversion | 367 |
| 8.2.5 | Weak Inversion | 378 |
| 8.2.6 | Moderate Inversion | 380 |
| 8.2.7 | General Models | 381 |
| 8.3 | A Medium-Frequency Small-Signal Model for the Intrinsic Part | 384 |
| 8.3.1 | Introduction | 384 |
| 8.3.2 | Intrinsic Capacitances | 385 |

| | | |
|-------------------|---|------------|
| 8.4 | Small-Signal Modeling for the Extrinsic Part | 405 |
| 8.5 | Noise | 410 |
| 8.5.1 | Introduction | 410 |
| 8.5.2 | White Noise | 414 |
| 8.5.3 | Flicker Noise | 422 |
| 8.5.4 | Small-Dimension Effects | 424 |
| 8.5.5 | Equivalent-Circuit Model | 425 |
| 8.6 | General Models | 426 |
| | References | 428 |
| | Problems | 438 |
| Chapter 9 | High-Frequency Small-Signal Models | 440 |
| 9.1 | Introduction | 440 |
| 9.2 | A Complete Quasi-Static Model | 440 |
| 9.2.1 | Complete Description of Capacitance Effects | 440 |
| 9.2.2 | Small-Signal Equivalent Circuit Topologies | 445 |
| 9.2.3 | Evaluation of Capacitances | 451 |
| 9.2.4 | Frequency Region of Validity | 458 |
| 9.3 | y-Parameter Models | 460 |
| 9.4 | Non-Quasi-Static Models | 467 |
| 9.4.1 | Introduction | 467 |
| 9.4.2 | A Non-Quasi-Static Strong-Inversion Model | 469 |
| 9.4.3 | Other Approximations and Higher-Order Models | 490 |
| 9.4.4 | Model Comparison | 491 |
| 9.5 | High-Frequency Noise | 492 |
| 9.6 | Considerations in MOSFET Modeling for RF Applications | 495 |
| | References | 504 |
| | Problems | 508 |
| Chapter 10 | MOSFET Modeling for Circuit Simulation | 513 |
| 10.1 | Introduction | 513 |
| 10.2 | Types of Models | 514 |
| 10.3 | Combining Several Effects into One Physical Model | 516 |
| 10.4 | Parameter Extraction | 519 |
| 10.5 | Accuracy | 528 |
| 10.6 | Properties of Good Models | 529 |
| 10.7 | Considerations and Choices | 530 |
| 10.7.1 | General Considerations | 530 |
| 10.7.2 | Considerations Related to Computer Implementation | 533 |
| 10.8 | Benchmark Tests | 534 |
| 10.9 | Nontechnical Considerations | 543 |
| | References | 543 |
| | Problems | 553 |

| | |
|------------|---|
| Appendices | |
| A | Energy Bands and Related Concepts 557 |
| B | Basic Laws of Electrostatics in One Dimension 566 |
| C | Charge Density, Electric Field, and Potential in the pn Junction 572 |
| D | Energy Band Diagrams for the Two-Terminal MOS Structure 574 |
| E | Charge Density, Electric Field, and Potential in the Two-Terminal MOS Structure 578 |
| F | General Analysis of the Two-Terminal MOS Structure 580 |
| G | Careful Definitions for the Limits of Moderate Inversion 584 |
| H | Energy Band Diagrams for the Three-Terminal MOS Structure 587 |
| I | General Analysis of the Three-Terminal MOS Structure 591 |
| J | Drain Current Formulation Using Quasi-Fermi Potentials 594 |
| K | Results of a Detailed Formulation for the Drain Current and Drain Small-Signal Conductance in the Saturation Region 598 |
| L | Evaluation of the Intrinsic Transient Source and Drain Currents 600 |
| M | Charges for the Accurate Strong-Inversion Model 603 |
| N | Quantities Used in the Derivation of the Non-Quasi-Static y -Parameter Model 606 |
| | Index 609 |

CHAPTER 1

SEMICONDUCTORS, JUNCTIONS, AND MOSFET OVERVIEW

1.1 INTRODUCTION

The discussion of MOS devices in this book will be based on an understanding of a few basic concepts. These concepts have been collected in this chapter. We begin with an introduction to semiconductors and the evaluation of mobile carrier concentrations in them. We then consider the mechanisms of current transport in semiconductors. We continue with a discussion of contacts between different materials and the electrostatic potentials established in such contacts. One special contact, the *pn* junction, is then discussed. The material here is meant primarily as a review, but has been written in such a way that it can be understood even with no prior exposure to semiconductor electronics. It has thus been kept as simple as possible. A more advanced and detailed treatment can be found in several textbooks.¹⁻¹²

The chapter concludes with a brief overview of the MOS transistor and of the chapters that follow.

1.2 SEMICONDUCTORS

Semiconductors derive their name from the fact that they can conduct current better than insulators, but not as well as conductors. The most widely used semiconductor material currently is silicon. The following discussion is focused on this material, but

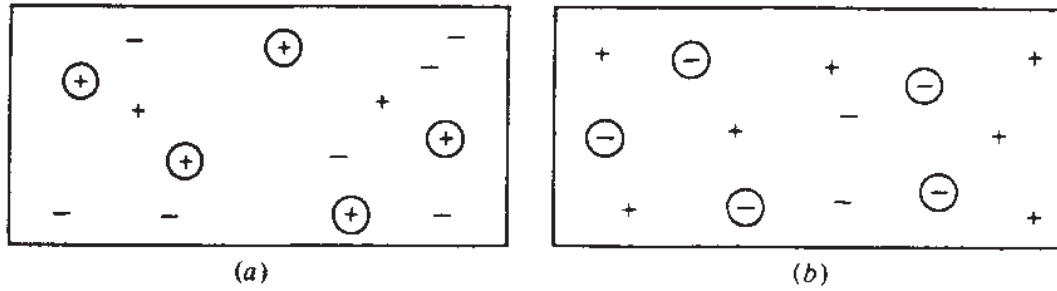
the qualitative arguments used are valid for other semiconductors as well. Throughout this book we assume that no illumination, no radiation, no mechanical stress, and no magnetic fields are present, and that all points of the semiconductor are at the same temperature (understood to be room temperature unless indicated otherwise). Until further notice we also assume that the semiconductor material under discussion is self-contained, with no externally applied voltage or current, and that the electric field is zero in its environment (the assumption of zero electric field will be relaxed later on in this section). Finally, we will assume that all of the above have been satisfied for a long time, so that conditions within the semiconductor have settled. The semiconductor is then said to be in equilibrium.

A pure (intrinsic) silicon crystal consists of an orderly three-dimensional array of atoms. This array is called the *crystal lattice* and contains approximately 5×10^{22} atoms/cm³. The atoms of the lattice are held together by cooperating electrons, called *valence* electrons, which form *bonds* between the atoms. At absolute zero temperature, all such electrons are firmly held in place, and the total negative charge of the electrons in each atom is canceled by an opposite positive charge contained in the atom's nucleus. At higher temperatures, the lattice vibrates due to thermal energy; this "thermal motion" manages to set some of the electrons loose from the parent atom. These become *free electrons*, in the sense that they are now free to move about the crystal; the name "free electrons" is used to distinguish them from the rest of the electrons that are still part of the bonds between atoms and are not free to wander away. If the motion of free electrons is coordinated, it can cause the flow of current. Since the atoms from which these electrons broke loose were electrically neutral originally, they are now left with a net positive charge.

Consider now two neighboring atoms, *A* and *B*, and assume that an electron was set free from *A*; now there is an electron vacancy in *A*, so *A* is overall positive. A valence electron associated with atom *B* can move and fill this vacancy, thus creating now a vacancy in *B*. Notice that this electron moved from one bond to another, i.e., did not become free. The result of this valence electron transfer is that now *A* is neutral, whereas *B* has acquired a net positive charge. A valence electron from an atom *C* near *B* can now move, fill the vacancy in *B*, thus making *B* neutral and *C* positive, etc. We see that this mechanism transports a positive charge from *A* to *B* to *C*.

Thus, we encounter *two* mechanisms for carrying charge around the semiconductor: (1) The motion of *free* electrons about the crystal lattice, each such electron carrying a negative charge, and (2) the motion of valence electrons from bond to bond, corresponding to a motion of "vacancies" (and associated positive charges) in the opposite direction. This second phenomenon can be described by modeling it as a motion of fictitious free particles, called *holes*, which carry a positive charge; each hole can be associated with one vacancy. If the charge of one electron is denoted by $-q$, then the charge of one hole is $+q$. As they wander around the lattice, a hole and a free electron can meet and annihilate each other; this is called *recombination*. The picture of holes and free electrons provided here is sufficient for our purposes. It should be noted, however, that this picture is only a simple *model* for what is actually a combination of very complex physical phenomena.

In the pure semiconductor we are discussing, since each hole is created by the breaking loose of one electron which becomes free, there is an equal number of holes

**FIGURE 1.1**

Free electrons, holes, and ionized dopant atoms in extrinsic silicon. (a) *n* type; (b) *p* type.

and free electrons. Thus, let the volume concentration of electrons and holes in the intrinsic material be denoted by n_i and p_i respectively. We will have

$$p_i = n_i \quad (1.2.1)$$

The symbol n_i is often used to denote either concentration and is referred to simply as the *intrinsic carrier concentration*. Its value corresponds to the equilibrium case where the rates of generation and recombination of electron-hole pairs are equal. At 300 K (kelvins) the value¹¹ of n_i for silicon is approximately $1.18 \times 10^{10} \text{ cm}^{-3}$ (other values are also in use). This means that roughly 12 electron-hole pairs are to be found in a cube $10 \text{ } \mu\text{m}$ on a side. Since there are $5 \times 10^{22} \text{ silicon atoms/cm}^3$, only about two out of every 10^{13} of these atoms contribute one electron-hole pair! As might be expected from our previous discussion, n_i increases at higher temperatures.[†]

It is possible to make the number of free electrons different from that of the holes by introducing foreign atoms in the silicon crystal. Such atoms are called *impurities*, and the process of introducing them is called *doping*; doped semiconductors are referred to as *extrinsic semiconductors*. If it is desired to enhance the free-electron population, these foreign *dopant* atoms are chosen so as to have available for bonding one electron *more* than the number needed for perfect bonding in the silicon structure. Thus, these atoms form bonds with neighboring silicon atoms by using up all their valence electrons except one. The latter is very loosely held to the parent atom, and at room temperature the thermal vibration of the crystal lattice is enough to set it free. In contrast to the case of the intrinsic (pure) semiconductor, the departure of this electron leaves all valence bonds intact; hence it does not cause a vacancy in them, and thus it does not leave behind a hole. However, since the dopant atom was originally neutral, it is now left with a net positive charge and is said to be *uncovered* or *ionized*. What we have described is shown in Fig. 1.1a. Each - sign represents one free electron. Each *circled* + sign represents a dopant atom which has lost one electron, and thus is left with a net positive charge. The circle is used to

[†]An approximate formula for n_i in silicon as a function of the absolute temperature T is $n_i = A_1 T^{3/2} \exp(-A_2/T)$, where $A_1 = 3.1 \times 10^{16} \text{ K}^{-3/2} \text{ cm}^{-3}$ and $A_2 = 7000 \text{ K}$.

indicate that this atom, being part of the crystal lattice, is *immobile* and therefore is not itself available for conduction. At room temperature practically all dopant atoms are ionized, and thus the number of free electrons created by such ionization is practically equal to the number of these atoms. As already mentioned, each such atom contributes one free electron without creating a hole. However, a few hole-electron pairs are created by the silicon atoms due to mechanisms already described for the intrinsic (pure) case; two such pairs are included in Fig. 1.1a (a hole is represented by an uncircled + sign).

Since the dopant atoms in Fig. 1.1a were chosen such as to “donate” one free electron to the silicon crystal, they are called *donors*. Donor materials commonly used to dope silicon are phosphorus, arsenic, and antimony. The donor atoms are introduced into the silicon in very minute amounts, but their concentration is usually chosen several orders of magnitude higher than n_i . For example, although a donor atom concentration of 5×10^{16} per cm^3 (written cm^{-3}) corresponds to one donor atom in every million silicon atoms, this concentration is more than a million times larger than the intrinsic carrier concentration at room temperature. The donor concentration will be assumed uniform unless stated otherwise. If all donor atoms are ionized, the number of free electrons they contribute is much higher than those contributed by the “intrinsic” mechanism discussed previously; hence, the concentration of free electrons, denoted by n_o , is approximately equal to the donor concentration, denoted by N_D :

$$n_o \approx N_D \quad (1.2.2)$$

With so many free electrons moving around the lattice, the chance of their encountering a hole in their way and filling it, or “recombining” with it, is significant; thus the concentration of holes, denoted by p_o , decreases compared to that in the intrinsic case. In fact, this chance of recombination is approximately proportional to n_o , and thus p_o decreases by the same factor that n_o has increased (compared to the intrinsic case). Hence, the product $n_o p_o$ remains the same as in the intrinsic case,¹ i.e., it is equal to n_i^2 as seen from (1.2.1). From this fact and (1.2.2) it follows that

$$p_o \approx \frac{n_i^2}{N_D} \quad (1.2.3)$$

If the doping concentration is very high (higher than about 10^{18} cm^{-3}), the above will not hold.^{1,5,10} Semiconductors with very high doping concentration are said to be *degenerate*. Also, the above two relations will not hold at very low temperatures, where the dopant atoms will not all be ionized, or at very high temperatures, where n_i rises to the point that the assumption $N_D \gg n_i$ is not valid. Whenever the above relations are used, it will be implied that none of these extreme situations is in effect.

Because in a donor-doped semiconductor n is larger than p , the free electrons are called the *majority carriers* and the holes the *minority carriers*. A semiconductor doped with donor impurities is said to be n type, because the majority carriers in it carry a *negative* charge.

Instead of increasing the mobile electron population as described above, it is possible to increase the *hole* population by introducing into pure silicon impurity atoms which have one valence electron *less* than the number needed for complete bonding with neighboring silicon atoms. Thus, when each such atom attempts to form bonds, it will be short of one valence electron. It can then “steal” such an electron from a neighboring silicon atom, a process which has two effects. First, since the impurity atom was electrically neutral originally, now that it has acquired an extra electron, it will have a net negative charge; this charge is associated with one specific atom and is thus immobile in the sense that it is not available for current conduction. Second, this stealing away one electron from a neighboring silicon atom left a valence electron vacancy in the latter, and thus created a hole; this hole can move around as in the case of the intrinsic (pure) crystal. Notice, however, that in contrast to the intrinsic case, the hole was created *without*, at the same time creating a free electron.

Since the impurity atoms have stolen or “accepted” one valence electron from the silicon lattice, they are called *acceptors*; typical acceptor materials used to dope silicon are boron, gallium, and indium. Fig. 1.1*b* illustrates the charges in a semiconductor doped with acceptor atoms. The + signs denote holes, the circled – signs denote ionized acceptor atoms (which are immobile), and the uncircled – signs denote free electrons (as in the case of donor-doped silicon, a few hole-free electron pairs are still created by silicon atoms due to the mechanisms described for the intrinsic case above). The total charge in Fig. 1.1*b* adds up to zero, indicating that the semiconductor is macroscopically neutral. Let the acceptor concentration (assumed uniform) be denoted by N_A , and assume that $N_A \gg n_i$ (a typical value of N_A is 10^{17} cm^{-3}). Assuming that practically all acceptor atoms are ionized, we will have, since each atom contributes one hole,

$$p_o \approx N_A \quad (1.2.4)$$

As in the case of donor-doped material, the product $n_o p_o$ remains equal to n_i^2 ; hence, we have

$$n_o \approx \frac{n_i^2}{N_A} \quad (1.2.5)$$

The above approximations will fail at extremely low or high temperatures, or if the doping concentration is extremely high, as explained for the case of donor doping. In an acceptor-doped semiconductor, the holes are the “majority carriers,” and the electrons the minority carriers. Since majority carriers carry a *positive* charge, semiconductors doped with acceptor impurities are said to be *p type*.

The above discussion has assumed zero electric field.[†] If the electric field within the semiconductor is not zero, the relations presented above will not be valid

[†]The term *electric field* is reserved here for macroscopic electric fields. Of course fields will always be present at the atomic level, e.g., between electrons and the nucleus in a given atom.

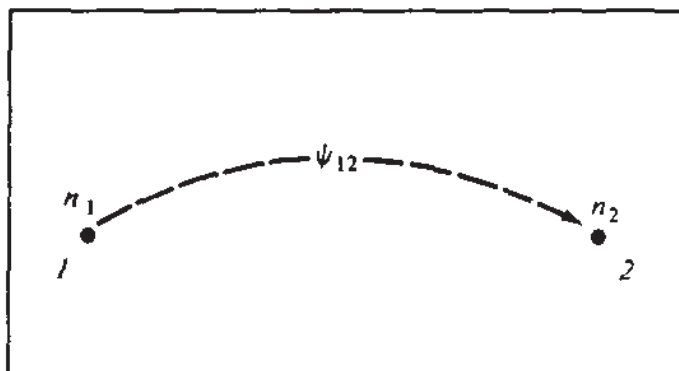


FIGURE 1.2

A piece of semiconductor material with different electron concentration at two points.

in general. Nevertheless, in certain of these cases the semiconductor is still said to be in equilibrium.^{1,6} We will encounter cases that fall into this category, which will be characterized by the lack of net energy exchange between the semiconductor and the external world and no net current flow; for example, this will be the case for a short-circuited pn junction, discussed in Sec. 1.5. In such cases, the following will continue to be true as before^{1,6} (with n and p denoting the electron and hole concentrations, respectively),

$$np = n_i^2 \quad (1.2.6)$$

whether the semiconductor is extrinsic or intrinsic. Note that in general the individual values of n and p will be *different* from the values of n_o and p_o found in the absence of an electric field. Thus, consider a region of semiconductor material (intrinsic, n type, or p type) in equilibrium. Assume that an electrostatic potential difference ψ_{12} exists between two points 1 and 2, as shown in Fig. 1.2.† Then it can be shown by using “energy band” concepts (Appendix A) that the electron concentrations n_1 and n_2 at points 1 and 2, respectively, are related to ψ_{12} by

$$\frac{n_1}{n_2} = e^{\psi_{12}/\phi_t} \quad (1.2.7)$$

where the quantity ϕ_t is the so-called “thermal voltage,” given by

†In this book, the potential of a point with respect to another will be indicated by an arrow pointing *from* the former point *to* the latter. If the potential is internal to a device, the arrow will be drawn with a broken line. Potentials between a device's external terminals will be indicated by solid-line arrows.

$$\boxed{\phi_i = \frac{kT}{q}} \quad (1.2.8)$$

with k the Boltzmann constant, q the magnitude of the electron charge, and T the absolute temperature. The values for the constants k and q are

$$k = 1.3809 \times 10^{-23} \frac{\text{C} \cdot \text{V}}{\text{K}} = 8.62 \times 10^{-5} \frac{\text{eV}}{\text{K}} \quad (1.2.9)$$

$$q = 1.602 \times 10^{-19} \text{ C} \quad (1.2.10)$$

From the above three equations, the value of ϕ_i at room temperature (300 K) is 0.0259 V.

The relation corresponding to (1.2.7) for holes is (again, assuming equilibrium; Appendix A),

$$\boxed{\frac{p_1}{p_2} = e^{\psi_{21}/\phi_i}} \quad (1.2.11)$$

Note that (1.2.7) and (1.2.11) together imply $n_1 p_1 = n_2 p_2$, just as would be the case in the absence of any potential difference in equilibrium.

Let us now consider the charge density (charge concentration per unit volume) in a semiconductor. In the general case, four entities can be responsible for its value: (1) holes, which contribute a charge density of $(+q)p$; (2) free electrons, with contribution $(-q)n$; (3) ionized donor atoms, with contribution $(+q)N_D$; and (4) ionized acceptor atoms, with contribution $(-q)N_A$. The total charge density, denoted by ρ , will be the sum of the individual contributions:

$$\boxed{\rho = q(p - n + N_D - N_A)} \quad (1.2.12)$$

Of course, in the cases already discussed, we will have $N_A = 0$ if only donor atoms are present, or $N_D = 0$ if only acceptors are present. There are cases, however, where both N_D and N_A can be nonzero. For example, if it is desired to convert part of a p -type region into an n -type region, one can introduce donor atoms at a concentration higher than the concentration of acceptor atoms. The region then becomes effectively n type, with an “effective” donor concentration of $N_D - N_A$. This process is encountered often in the fabrication of semiconductor devices.

In the presence of electric fields, the charge density ρ can vary from point to point. Assuming equilibrium, n and p in (1.2.12) must be such that, for any two points, they are related to the electrostatic potential by (1.2.7) and (1.2.11), which are consequences of semiconductor properties. In addition to these relations, the total charge density ρ must satisfy *Poisson's equation*, which is a general relation in

electrostatics and is not restricted to semiconductors. Let us consider the “one-dimensional case” where ρ and the electrostatic potential ψ (taken with respect to some arbitrary reference) vary only vertically (along the y direction)[†]; then Poisson’s equation is (Appendix B)

$$\boxed{\frac{d^2\psi}{dy^2} = -\frac{\rho(y)}{\epsilon_s}} \quad (1.2.13)$$

where ϵ_s is the “permittivity” of the material, given by

$$\epsilon_s = k_s \epsilon_0 \quad (1.2.14)$$

with ϵ_0 the permittivity of free space (8.854×10^{-14} F/cm) and k_s the dielectric constant of the material. For silicon we have approximately $k_s = 11.8$, corresponding to $\epsilon_s = 1.04 \times 10^{-12}$ F/cm.

Our analysis of semiconductor devices in equilibrium will be based on the simple relations we have provided in this section. Several examples of the application of these relations will be found throughout this book. For the nonequilibrium cases of interest to us, a simple modification will be necessary, as will be seen.

1.3 CONDUCTION

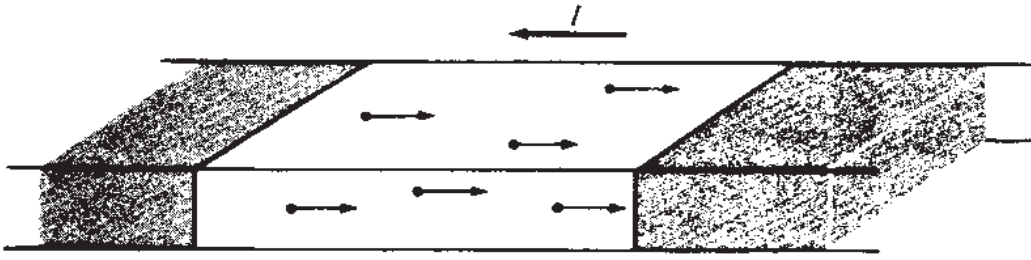
1.3.1 Transit Time

The notion of “transit time” will be used in our discussion of MOS transistors in subsequent chapters. This notion is general and independent of the mechanisms via which current conduction takes place. Thus, it is introduced first before such mechanisms are discussed.

Consider the piece of material shown between the shaded regions in Fig. 1.3. Assume a steady flow of current in one direction, wholly due to free electrons, has been established. Also, assume that *no* recombination of electrons with holes takes place. Electrons are constantly being supplied from the left side and taken out from the right side at the same fixed rate. At any given instant, then, the magnitude of the total electron charge that happens to be inside the piece is fixed; let $|Q|$ denote its value. We will make the simplifying assumption that it takes each electron the same amount of time to travel the length of the piece; this time will be called the *transit time* and will be denoted by τ .

Let us consider the free electrons found inside the piece at a given instant and follow their motion. After the lapse of one transit time τ even those initially at the far left will have exited through the right end. Therefore all the electrons inside the piece at the initial instant of observation will have exited through the right-hand end also

[†]This direction is chosen in anticipation of analyses to follow later.

**FIGURE 1.3**

A piece of material with electrons flowing toward the right.

and will have been replaced by new electrons coming from the left. Thus, a *negative* charge of magnitude $|Q|$ exits from the right-hand end in time τ ; this corresponds to a current I in the direction shown, given by

$$I = \frac{|Q|}{\tau} \quad (1.3.1)$$

Note that in the argument leading to (1.3.1) we did not make any assumption as to the detailed mechanisms through which the current is conducted, the distribution of charge inside the material, the presence or absence of electric fields, or the constancy of electron velocity along the current path. Thus, (1.3.1) is quite general.

In the above discussion, we made the convenient assumption that all electrons spend the same amount of time τ in traveling the length of the material. However, it is customary to relax this assumption and use

$$\tau = \frac{|Q|}{I} \quad (1.3.2)$$

as the *definition* of transit time, with $|Q|$ and I as defined above. An obviously similar definition can be given in the case of conduction caused by holes. Expressions for τ will be found for each of two mechanisms of current conduction, which are considered in the next two subsections.

1.3.2 Drift

Let us consider a piece of semiconductor with no external field applied to it. A random “thermal” motion is exhibited by the holes and electrons in all directions; however, on the average these random charge movements cancel out and there is no net current produced. If an electric field is now applied (for example, by connecting the semiconductor piece across the terminals of a battery), it will exert forces on the charged particles. Thus there will be a net movement along the field lines which can be observed macroscopically as an electric current. This phenomenon is known as *drift*; it would not occur if the particles were not charged.

The movement of electrons and holes during drift is quite complicated since these carriers interact with their environment, for example, with the thermal vibrations

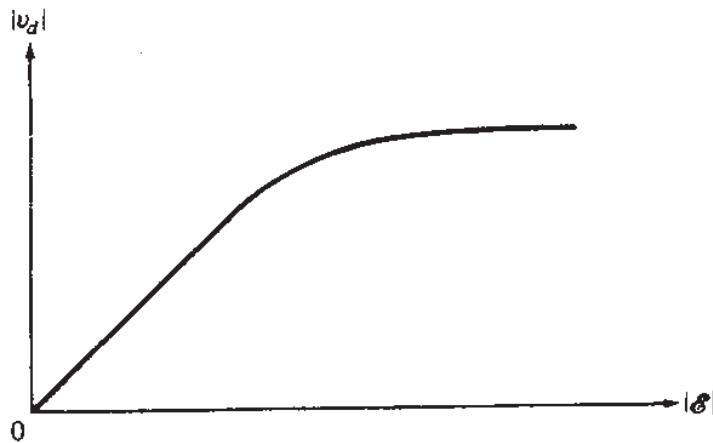


FIGURE 1.4
Magnitude of drift velocity versus magnitude of electric field.

of the lattice and with ionized impurity atoms.[†] Such interactions are referred to as *scattering*. For example, as an electron accelerates in the direction dictated by the electric field, it can be scattered and lose some of its energy to its environment, then it accelerates again, etc. Because of this, and because charge is transported by these carriers in discrete amounts, a minute “noise” fluctuation exists in the externally observed current. For the present such fluctuations will be ignored and attention will be focused on the average current value I , which is nonzero because of the net movement caused by the nonzero electric field. The quantity I can be calculated from the carrier’s *average* velocity, called *drift velocity* and denoted by v_d . For a given electric field, v_d depends on the type of semiconductor, the type and concentration of doping, the temperature, and the type of carrier.⁵ For silicon, the magnitude of the drift velocity depends on the magnitude of the electric field \mathcal{E} as shown qualitatively in Fig. 1.4. For high electric fields, the loss of energy of the carriers to the lattice becomes more effective and eventually *velocity saturation* is reached as shown. The maximum velocity value is essentially independent of doping concentration, and for silicon at room temperature is of the order of 10^7 cm/s for both electrons and holes. It is reached at fields roughly above 3×10^4 V/cm for electrons and 10^5 V/cm for holes. Writing these values as 3 V/ μ m and 10 V/ μ m, respectively, we see that a few volts across a length of 1 μ m are enough to cause velocity saturation.

We now consider a uniformly doped n -type semiconductor bar with dimensions a , b , c as shown in Fig. 1.5, and a voltage V applied across it. In the following we will neglect the contribution of hole movement to the electric current since the holes, being the minority carriers in this case, are much fewer than the electrons. Electrons move in a complicated fashion with an *average* velocity of v_d , as discussed above. We can calculate the resulting current by considering a simpler, hypothetical picture

[†]Our present discussion deals only with drift in the *bulk* of a semiconductor. Additional scattering mechanisms are present at the semiconductor surface in an MOS transistor; these are considered in Chap. 4.

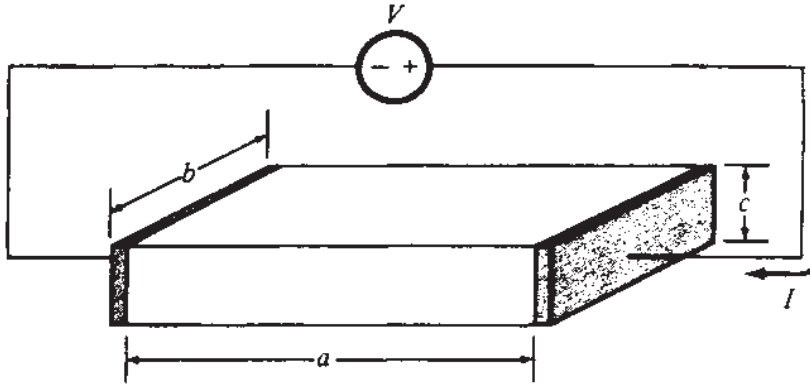


FIGURE 1.5
An n -type semiconductor bar with uniform electron concentration under external bias.

in which the velocity of all electrons is constant and equal to v_d . The time it takes for an electron to travel the length of the bar is

$$\tau = \frac{a}{|v_d|} \quad (1.3.3)$$

The magnitude $|Q|$ of the total free electron charge found inside the bar at a given time instant is given by the charge magnitude of a single electron q times the total number of electrons; the latter is given by the volume of the bar times the free electron concentration per unit volume, n . Thus

$$|Q| = nq(abc) \quad (1.3.4)$$

Using this with (1.3.1) and (1.3.3) gives

$$I = \frac{nq(abc)}{\tau} \quad (1.3.5a)$$

$$= nq(bc)|v_d| \quad (1.3.5b)$$

This will now be put in a different form that will prove useful later in the discussion of transistors. The area of the bar's top surface in Fig. 1.5 is (ab) . Thus $|Q|/(ab)$ is the *magnitude of the charge per unit area*.[†] This quantity will be denoted by $|Q'|$. Using (1.3.4) we have

$$|Q'| = \frac{|Q|}{ab} = nqc \quad (1.3.6)$$

[†]The use of this quotient is adequate because the charge is uniformly distributed horizontally; otherwise, the charge per unit area would have to be defined differentially.

Substituting this in (1.3.5b) gives

$$I = b|Q'| |v_d| \quad (1.3.7)$$

THE CASE OF LOW ELECTRIC FIELDS. The above relations assume special, very useful forms in the case of low electric fields. For silicon, “low” means roughly less than 3×10^3 V/cm (i.e., 0.3 V/ μ m) for electrons, and less than 6×10^3 V/cm (i.e., 0.6 V/ μ m) for holes. For such fields, $|v_d|$ is proportional to $|E|$, as suggested by the bottom part of the curve in Fig. 1.4. The constant of proportionality is called *mobility* and is denoted by μ_B . The subscript *B* is used to emphasize that this mobility characterizes the “bulk” of the semiconductor. This is needed to distinguish it from a “surface” mobility which will be introduced in Chap. 4. We thus have

$$|v_d| = \mu_B |E| \quad (1.3.8)$$

Electron and hole mobilities for silicon at room temperature are shown in Fig. 1.6 versus doping concentration.¹³ Mobilities decrease with temperature under common conditions.^{5,13}

The magnitude of the electric field in Fig. 1.5 is given by:

$$|E| = \frac{V}{a} \quad (1.3.9)$$

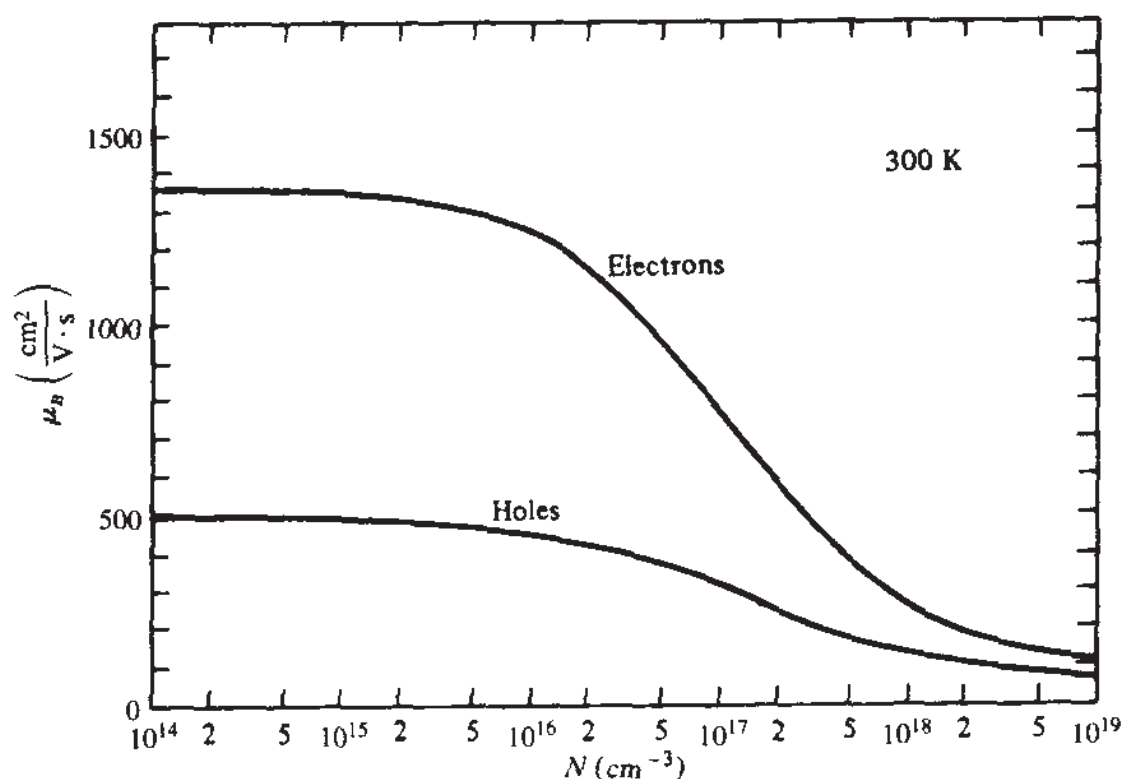


FIGURE 1.6

Electron and hole *bulk* mobility in silicon at 300 K vs. doping concentration.¹³

With the use of this in (1.3.8) and of the result in (1.3.3) we obtain, for the low-field transit time,

$$\tau = \frac{a^2}{\mu_B V} \quad (1.3.10)$$

The presence of the a^2 term reflects the fact that, for fixed μ_B and V , increasing the bar's length increases the transit time for two reasons: (1) The distance the electrons must travel becomes larger and (2) the magnitude of the electric field becomes smaller, which slows the electrons down.

Using the above equation in (1.3.5a) we obtain:

$$I = \mu_B n q \frac{bc}{a} V \quad (1.3.11)$$

which is nothing but Ohm's law.† This equation can be written in the form $I = GV$, where G is called the conductance and is given by

$$G = \sigma \frac{bc}{a} \quad (1.3.12)$$

where σ is the *conductivity* given by

$$\sigma = \mu_B n q \quad (1.3.13)$$

The *inverse* of the conductivity, called the *resistivity*, is also used.

Using the above two equations and (1.3.6), we can express the conductance as follows:

$$G = \mu_B |Q'| \frac{b}{a} \quad (1.3.14)$$

The I - V relation is thus

$$I = \mu_B |Q'| \frac{b}{a} V \quad (1.3.15)$$

†We have assumed a uniform n throughout the bar. However, we will encounter some cases where this assumption is not valid. If n varies along the x axis, one must instead consider a vertical "slice" of length Δx centered at x . Let $\Delta\psi$ be the electrostatic potential drop across the slice, defined from its right side to its left side; letting Δx approach zero, we obtain an equation in which the role of V/a in (1.3.11) is played by $d\psi/dx$:

$$I = \mu_B n(x) q b c \frac{d\psi}{dx} \quad (1.3.11a)$$

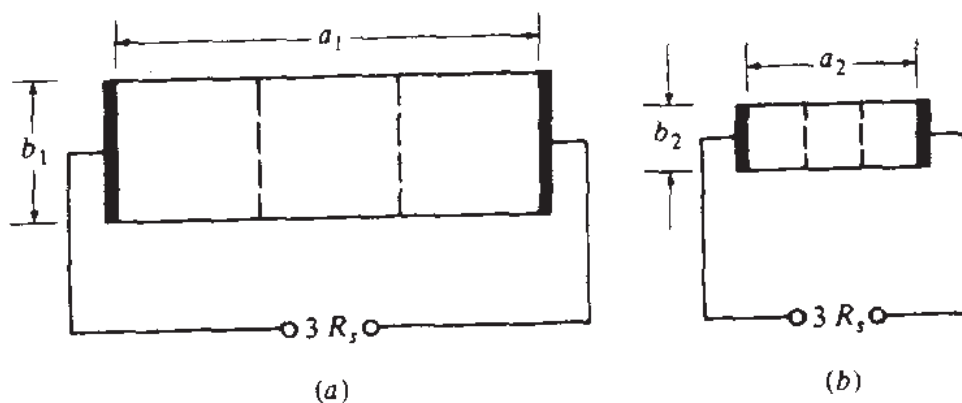


FIGURE 1.7

Top view of two bars, each with three "squares" in the path of the current.

As seen in the above two equations, the thickness c of the bar does not enter directly; all that counts is the value of the mobile charge concentration per unit area, and the dimensions b and a . In fact it is easy to show that *the above two results are valid even if the electron concentration is nonuniform in the vertical direction, as long as there is uniformity horizontally.*[†]

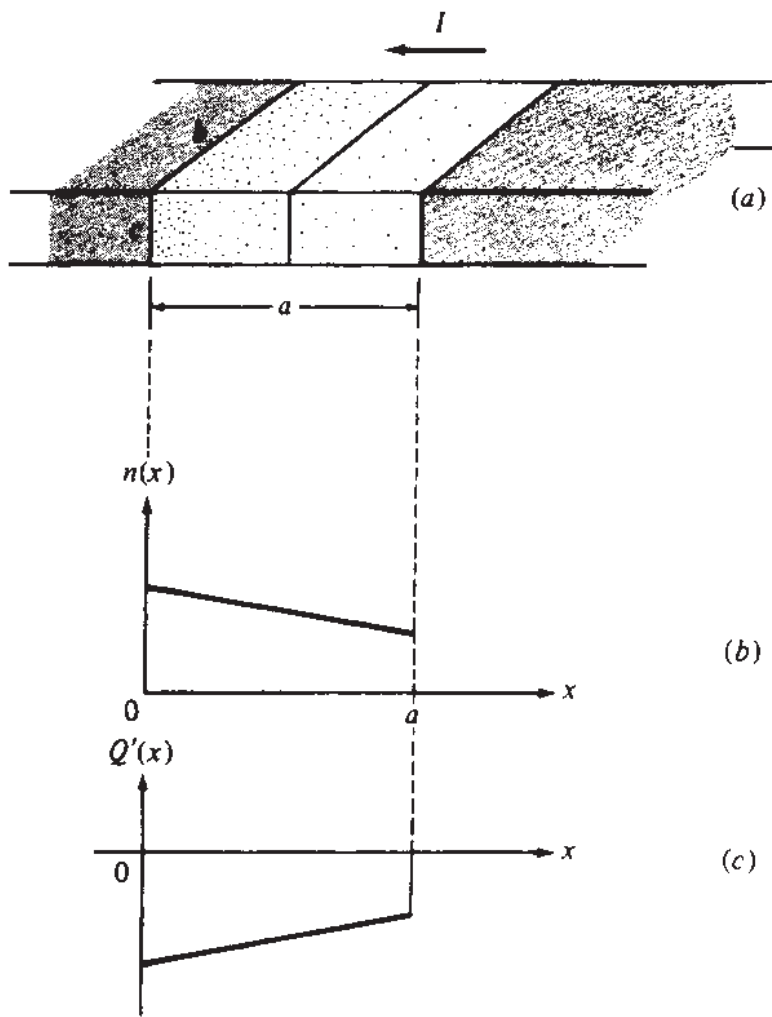
The resistance of the bar $R = 1/G$ is, from (1.3.14),

$$R = R_s \frac{a}{b} \quad (1.3.16)$$

where $R_s = (\mu_B |Q'|)^{-1}$ is called the *sheet resistance*. When $a = b$, i.e., when the bar as seen from the *top* is square, $R = R_s$. Thus R_s is simply the bar's "resistance per square"; R will be given by R_s times the number of squares in the path of the current. This is illustrated in Fig. 1.7. Both bars are assumed to be made out of the same material, and with the same Q' . Since both have a total of $a/b = 3$ squares in the path of the current, they have the same resistance, equal to $3R_s$. Sheet resistance is commonly expressed in "ohms per square."

Results analogous to the ones in this section can be given in the case of hole conduction. If holes were present in Fig. 1.5, they would move in a direction opposite to that of electrons (i.e., from right to left in the semiconductor). Since the charge associated with holes is positive, the resulting contribution to the total current would be in the direction shown, i.e., in the *same* direction as that for the current owing to electrons. One could then define a conductivity σ for use in (1.3.12), which would be the *sum* of the electron and hole conductivities.

[†]To show this, consider the bar as consisting of thin, parallel horizontal slices within each of which n is constant. Let ΔG and $\Delta Q'$ be the conductance and charge per unit area, respectively, for one slice. Then, from the above results $\Delta G = \mu_B |\Delta Q'| (b/a)$. Note that for the combination of all slices in parallel the total G and Q' are the sum of the conductances and the charges per unit area, respectively, for each slice. Letting ΔG and $\Delta Q'$ become differentials and integrating the resulting equation, we obtain (1.3.14) again. Equation (1.3.15) can be derived in a similar manner.

**FIGURE 1.8**

(a) A semiconductor bar with nonuniform electron concentration along its length; (b) the electron concentration in (a) for a special case of interest; (c) charge per unit area corresponding to (b).

1.3.3 Diffusion

Drift is only one of two major mechanisms responsible for the flow of electric current in semiconductors. The other mechanism, known as *diffusion*, occurs whenever particles are not distributed uniformly over space, i.e., when there exist “concentration gradients”; then the random motion of the particles tends to make them spread out from regions of high concentration to regions of low concentration. Notice that this phenomenon is not due to electric fields and can thus occur independently of whether or not the particles are charged; for example, particles of smoke exhibit such diffusion. However if the particles *are* charged, as are electrons and holes, diffusion gives rise to movement of charge and thus to electric current.

We consider the origin of diffusion with the help of Fig. 1.8a. Here a piece of semiconductor of rectangular cross section with width b and thickness c is assumed to contain electrons distributed uniformly across any vertical plane (such as the one shown near the middle), but *nonuniformly* along the length. No holes are assumed to be present for now. We assume that the semiconductor communicates with the external world from left and right, so that a fixed electron distribution can be maintained.

This distribution can be represented by a fixed plot of n vs. x . For now we will assume this plot to be a straight line as shown in Fig. 1.8b.

Electrons are found on both sides of the plane shown near the middle in Fig. 1.8a, and they exhibit a random thermal motion; however, in a parallelepiped of very small length Δx immediately to the left of the plane, more electrons are to be found than those contained in an identical parallelepiped immediately to the right of the plane. Hence one can expect that in a given time interval more electrons will cross the plane moving from left to right, rather than from right to left. This corresponds to a net flow of electrons from left to right. Since each carries a negative charge, the current defined in the direction shown in the figure will be positive. This current is proportional to the magnitude q of the charge carried by each electron and the cross-sectional area bc . Also, the more negative the slope of n with x , the more positive the current will be; in fact it can be shown that I is proportional to $(-dn/dx)$. We have¹⁻¹¹

$$I = Dq(bc)\left(-\frac{dn}{dx}\right) \quad (1.3.17)$$

where D is a constant of proportionality, called the *diffusion constant*. This constant is related to the mobility μ_B by the so-called Einstein relationship¹⁻¹¹ (Prob. 1.11):

$$D = \mu_B \phi_t \quad (1.3.18)$$

where ϕ_t is the thermal voltage given by (1.2.8). Analogous relations hold for the diffusion of holes; of course the values for the diffusion constants of holes and electrons are different, corresponding to their different mobilities. Note that if the particles shown in Fig. 1.8a were holes, and their concentrations were decreasing towards the right, there would be a hole movement from left to right. Since the charge associated with holes is positive, this movement would correspond to a current in the opposite direction from that shown.

If the plot of n vs. x is not a straight line, dn/dx will vary with position and so will the diffusion current. The latter must now be written as $I(x)$, and will still be given by the right-hand side of (1.3.17). Since, in steady state, the *total* current must be the same at any point x , there must exist in this case a drift current component, varying with position in such a way that the total current (which, as can be shown, will be given by the superposition of diffusion plus drift) is independent of x . In general, current flow in semiconductors is the result of both drift and diffusion; for example, both mechanisms are encountered in a pn junction. Often, one of the two mechanisms dominates. In the rest of this section, we assume for simplicity that only diffusion current is present, and thus the plot of n vs. x is a straight line.

Consider a thin vertical slice of the material in Fig. 1.8a of volume $bc \Delta x$ centered around a point at x . The charge in this slice is $(-q)n(x)bc \Delta x$, where the $-$ sign corresponds to the negative electron charge. Dividing the slice charge by the area of the slice $b \Delta x$ as seen from the top, and letting Δx go to zero, we obtain the *charge per unit area*, which here is a function of x :

$$Q'(x) = (-q)cn(x) \quad (1.3.19)$$

This quantity is shown vs. x in Fig. 1.8c. Using (1.3.19) and (1.3.18) in (1.3.17), we obtain

$$I = \mu_B \phi_i b \frac{dQ'(x)}{dx} \quad (1.3.20)$$

Consider now a more general case in which we allow the electron concentration to be *nonuniform in the vertical direction*. Assume that on a horizontal plane at any given depth, the plot of n vs. x is still a straight line (we still assume that n is uniform in the direction perpendicular to the plane of the paper in Fig. 1.8). Since n is now nonuniform vertically, electrons would have a tendency to diffuse in the vertical direction, but we will assume that an appropriate externally applied vertical electric field prevents them from doing so. In other words, we assume that the electrons still move parallel to the x direction, so that *laminar flow is maintained*. Then one can easily show that (1.3.20) will still be valid.[†]

Since the plot of $Q'(x)$ vs. x is a straight line, its slope will be

$$\frac{dQ'}{dx} = \frac{Q'(a) - Q'(0)}{a} \quad (1.3.21)$$

Use of this in (1.3.20) gives:

$$I = \mu_B \phi_i b [Q'(a) - Q'(0)] \quad (1.3.22)$$

Since (1.3.20) is valid even if n is a function of both x and the vertical dimension under the assumptions following (1.3.20), then (1.3.22) will also be valid.

The transit time τ can now be easily calculated. It is easy to show that the total charge of the electrons found at any instant inside the bar will be (Prob. 1.3)

$$Q = ab \frac{Q'(a) + Q'(0)}{2} \quad (1.3.23)$$

Thus the transit time will be, from (1.3.2),

$$\tau = \frac{a^2}{\mu_B (2\phi_i)} \frac{Q'(0) + Q'(a)}{Q'(0) - Q'(a)} \quad (1.3.24)$$

[†]To show this, consider the material as consisting of thin, parallel horizontal slices within each of which n does not vary vertically. If ΔI and $\Delta Q'(x)$ are the current and charge per unit area for one such slice, we will have $\Delta I = \mu \phi_i b d[\Delta Q'(x)]/dx$. The total current I will be the sum of the individual ΔI , and in this summation the $\Delta Q'(x)$ can be grouped together to give the total charge per unit area of the material at x , $Q'(x)$. More precisely, letting ΔI and $\Delta Q'(x)$ become differentials and integrating, we obtain (1.3.20) again (Prob. 1.4).

In this book we will often be interested in the case where a “perfect sink” exists at $x = a$, reducing the electron concentration to 0 there; this means

$$Q'(a) = 0 \quad (1.3.25)$$

For this special case (1.3.24) becomes

$$\tau = \frac{a^2}{\mu_B(2\phi_i)} \quad (1.3.26)$$

Note that this form is very similar to that for drift in (1.3.10). However, in contrast to the drift case, where τ can be made small by applying a large V , here τ is fixed at a comparatively large value, due to the fixed, small value of $2\phi_i$ (about 52 mV at 300 K).

1.4 CONTACT POTENTIALS

Consider the junction of two different materials J_1 and J_2 with no external bias, as shown in Fig. 1.9a; either material can be a semiconductor or a metal. When the two materials are brought together, at first carriers move from one to the other because

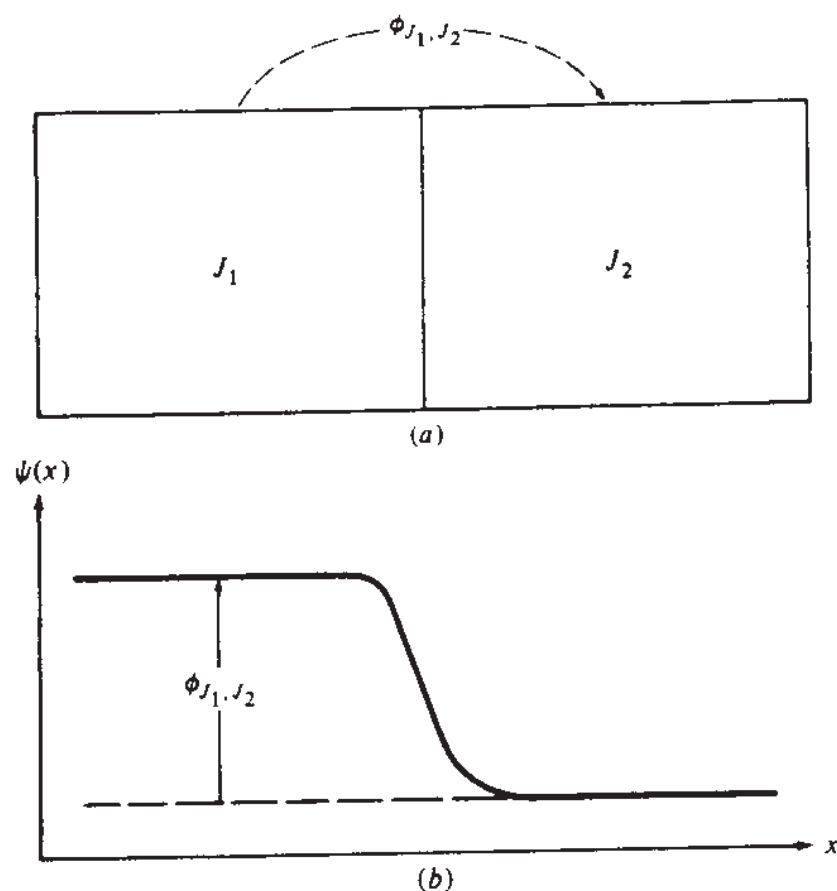


FIGURE 1.9
(a) Two different materials in contact (either can be a metal or a semiconductor); (b) potential vs. distance for (a).

the energy of these carriers is in general different in J_1 and J_2 , and no opposing field exists in the initially neutral materials. However, as each charged carrier crosses the junction, it leaves behind a net charge of the opposite polarity, and an electric field is thus established in the vicinity of the junction, which tends to inhibit the movement of carriers. For example, if an electron crosses from J_1 to J_2 , it leaves a positive charge in J_1 ; the contribution of this charge to the electric field is in such a direction as to attract the electron back into J_1 . Eventually, the field intensity increases to the point that it counteracts the tendency of carriers to cross the junction, and a balance is achieved such that there is no net carrier movement. An electrostatic potential change is then encountered in going from one material, through the junction, to the other material. If $\psi(x)$ is the potential at position x , it can be of the form shown in Fig. 1.9b (arbitrary references are assumed for both ψ and x). Depending on the two materials, the potential change can occur on both sides of the junction, as in the case shown, or mostly on one of them. The total potential drop in going from J_1 to J_2 is called the *contact potential of material J_1 to material J_2* , and will be denoted by ϕ_{J_1, J_2} . The concept of contact potentials will be instrumental in deriving the basic equations of MOS devices, without having to introduce the use of energy bands; hence, the use of contact potentials is discussed in considerable detail in this section. Readers interested in a treatment using energy bands are referred to Appendix A.

As an example consider the junction of a material J to intrinsic silicon, as in Fig. 1.10. Let ϕ_J denote the contact potential of material J to intrinsic silicon as shown. Depending on the material J , the value of ϕ_J will vary. For various materials it can be indirectly evaluated⁵ and is given in Table 1.1.

If J is a metal, the contact potential ϕ_J has a specific value as shown in the table. However, if J is a semiconductor, ϕ_J will depend not only on which semiconductor it is but also on its impurity type (donor or acceptor) and concentration. We will now focus on the case in which J is extrinsic silicon; the value of ϕ_J in this case is given by the *negative* of a quantity denoted by ϕ_F , as seen in the table. ϕ_F is a very important quantity characterizing the semiconductor which will be defined, for our purposes, as follows†:

The *Fermi potential* ϕ_F of extrinsic silicon is the contact potential that would be developed between intrinsic silicon and the extrinsic silicon if the two were

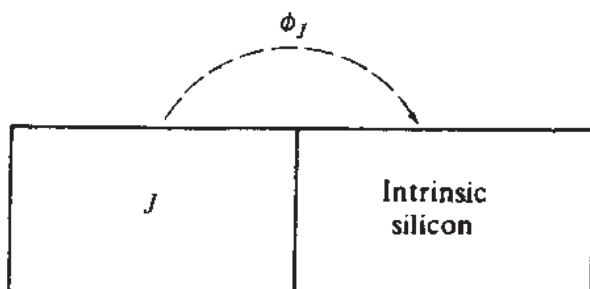


FIGURE 1.10

Junction of a material J and intrinsic silicon; definition of the characteristic potential ϕ_J .

†A formal definition of ϕ_F using energy band concepts is given in Appendix A.

TABLE 1.1
Approximate contact potential of
materials to intrinsic silicon (V)†

| Material | ϕ_J |
|------------------------------|-----------|
| Ag | -0.4 |
| Au | -0.3 |
| Cu | 0.0 |
| Ni | +0.15 |
| Al | +0.6 |
| Mg | +1.35 |
| Degenerate p^+ polysilicon | -0.56 |
| Degenerate n^+ polysilicon | +0.56 |
| Extrinsic Si | $-\phi_F$ |
| Intrinsic Si | 0 |

†The values given for metals and polycrystalline silicon are only approximate. Accurate values are difficult to obtain, and, in fact, different measuring techniques can yield values which differ by a large fraction of 1 V.

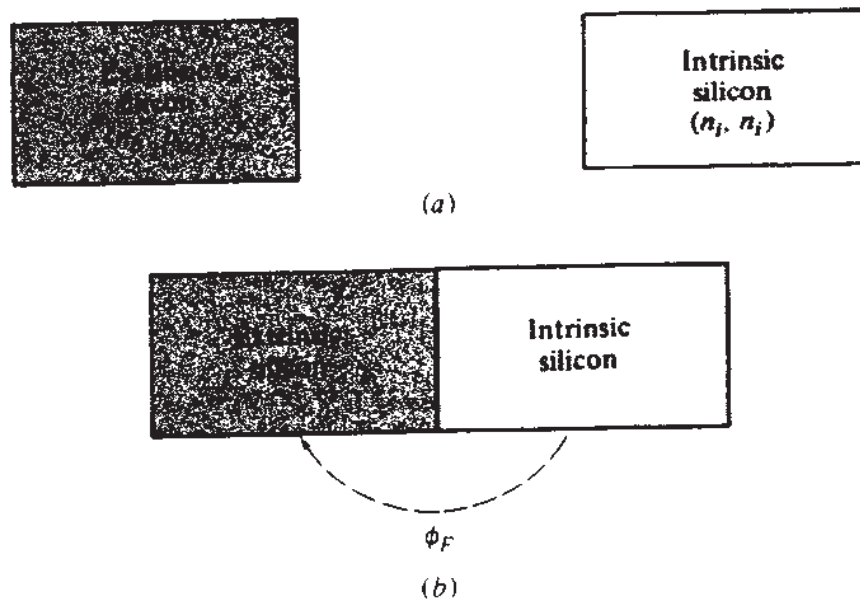


FIGURE 1.11

(a) Extrinsic and intrinsic silicon (symbols in parentheses indicate electron and hole concentrations);
 (b) the materials in (a) in contact with each other, and the corresponding contact potential.

brought in contact and thermal equilibrium were established. This contact potential is defined as that of intrinsic to extrinsic silicon†; see Fig. 1.11.

As one travels from the intrinsic side to the extrinsic side in Fig. 1.11b, there will be a transition region near the junction where the electrostatic potential will be changing, and the total change will be ϕ_F . At points clearly outside this transition

†Analogous definitions can be given for any semiconductor.

region, the electron and hole concentrations remain at the values they had before contact was made, i.e., they are as in Fig. 1.11a. Let 1 and 2 denote two such points in Fig. 1.11b, one in each side of the junction. We assume that there is no external bias and the system is in equilibrium. We can then apply (1.2.7) and (1.2.11) to obtain

$$\phi_F = \phi_i \ln \frac{n_i}{n_o} \quad (1.4.1a)$$

$$\phi_F = \phi_i \ln \frac{p_o}{n_i} \quad (1.4.1b)$$

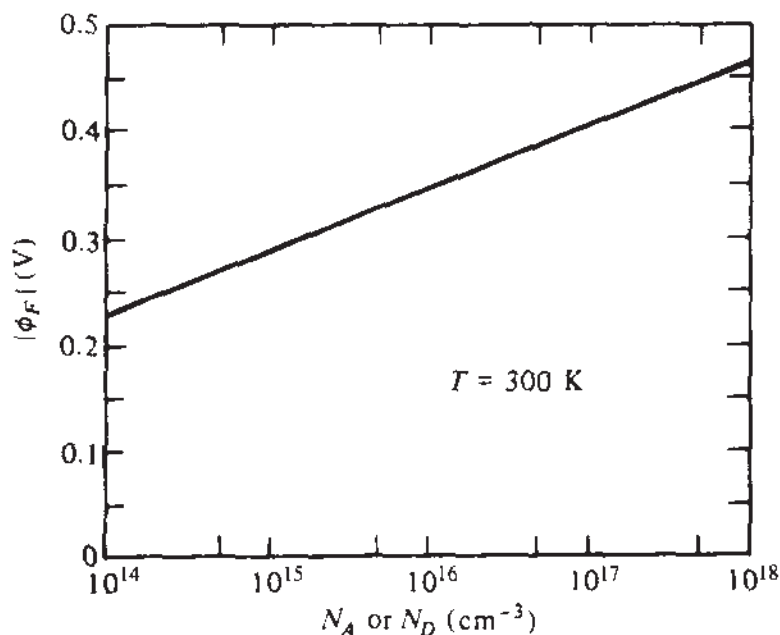
Let us now consider qualitatively the Fermi potential of *p*-type silicon. When it is joined to intrinsic silicon, one expects that a net number of electrons will diffuse from the intrinsic silicon to *p*-type silicon as electrons are in larger supply in the former, and holes will diffuse from the *p*-type side to the intrinsic side. Therefore, there will be a net positive charge on the intrinsic side and a net negative charge on the *p*-type side; the potential will be higher on the intrinsic side. Hence ϕ_F , which is the contact potential of the intrinsic side to the *p*-type side, will be positive. Now the higher the impurity concentration N_A on the *p* side, the more the holes and the larger the tendency to diffuse. Thus a larger field needs to be established to inhibit their movement at balance. This causes ϕ_F to be higher for higher N_A . Using (1.2.4) in (1.4.1b), we obtain

$$\phi_F \approx +\phi_i \ln \frac{N_A}{n_i} \quad p\text{-type material} \quad (1.4.2)$$

Similar qualitative considerations for *n*-type silicon lead to the conclusion that ϕ_F should be negative in that case, and that $|\phi_F|$ should increase with the donor concentration N_D . We can use (1.2.2) in (1.4.1a) to obtain

$$\phi_F \approx -\phi_i \ln \frac{N_D}{n_i} \quad n\text{-type material} \quad (1.4.3)$$

It should be mentioned here that, although we have discussed ϕ_F in terms of a contact potential, it is a quantity which characterizes the material by itself. Even if the material is self-contained (not in contact with anything else), to its free carrier concentration n (or p) there corresponds uniquely a value of ϕ_F at a given temperature. This value is given by (1.4.1) or, under the assumptions made in Sec. 1.2, by (1.4.2) or (1.4.3). Thus, for example, one can talk of a certain *p*-type silicon with $N_A = 10^{17} \text{ cm}^{-3}$ or, equivalently, of *p*-type silicon with $\phi_F = 0.413 \text{ V}$ at 300 K. Values for ϕ_F obtained from (1.4.2) and (1.4.3) at 300 K are plotted vs. doping concentration in Fig. 1.12. For degenerate materials, ϕ_F is difficult to evaluate. A common approximation is 0.56 V for *p*-type Si, and -0.56 V for *n*-type Si, at 300 K (Appendix A). The same estimates are used for degenerate polycrystalline silicon (often referred to simply as *polysilicon* or *poly*), a widely used material in modern transistors (Sec. 1.6).

**FIGURE 1.12**

Magnitude of Fermi potential vs. substrate doping concentration for silicon at room temperature.

We will now show that Table 1.1 is useful not only for calculating potentials in contacts with intrinsic silicon but also for calculating contact potentials for any two materials in that table, e.g., the contact potential of Au to Al. To see this, consider again the junction of two arbitrary materials J_1 and J_2 , as in Fig. 1.9a. The value of the contact potential, ϕ_{J_1, J_2} is, of course, independent of intrinsic silicon, since the latter appears nowhere in the figure. However, it is very easy to express the unknown ϕ_{J_1, J_2} in terms of ϕ_{J_1} and ϕ_{J_2} , which are known and tabulated in Table 1.1. To see this, suppose that we place the system of Fig. 1.9a in a loop with intrinsic silicon, as shown in Fig. 1.13. By equating the sum of the potential drops around the loop to zero, we get:

$$\phi_{J_1, J_2} = \phi_{J_1} - \phi_{J_2} \quad (1.4.4)$$

Therefore, intrinsic silicon is for us a *reference* material since, by knowing its contact potential to other materials, we can get the contact potential between any two materials as well. It is clear that we could have chosen any other material as our reference in Fig. 1.10. We could have then constructed a table for that material similar to Table 1.1, and (1.4.4) would be valid again for ϕ_{J_1} and ϕ_{J_2} obtained from the new table. Because we will be working with silicon devices, our earlier choice of intrinsic silicon as a reference will prove convenient.†

†In energy band treatments, contact potentials are often expressed in terms of the “work functions” of the two materials, work function W_J being a quantity characterizing material J and related to the energy required to remove an electron from the material to the vacuum. According to such treatments, ϕ_{J_1, J_2} in Fig. 1.9a is $(W_{J_2} - W_{J_1})/q$, where q is the magnitude of the electron charge (Appendix A). Work functions are not easy to measure accurately, and, in fact, have to be “modified” for use in MOS device modeling. We have avoided complications in our discussion by changing our reference from “vacuum,” which has little to do with the devices we will be dealing with, to something more relevant. If the reader prefers to use work functions, no confusion should arise; the difference in Eq. (1.4.4) is still correct if the symbol ϕ_J is defined to be $-W_J/q$.

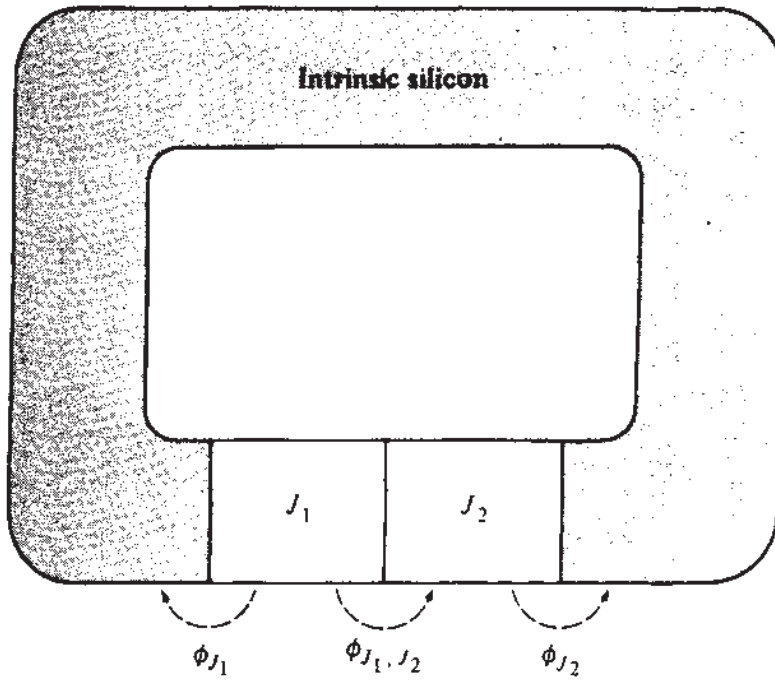


FIGURE 1.13

Two materials in a loop with intrinsic silicon.

Example 1.1. Calculate the contact potential of aluminum to p -type silicon, with $N_A = 10^{15} \text{ cm}^{-3}$, at $T = 300 \text{ K}$. From Table 1.1 and (1.4.2), and if P_1 denotes the above p -type material, we have

$$\phi_{P_1} = -0.0259 \text{ V} \ln \frac{10^{15}}{1.18 \times 10^{10}} = -0.29 \text{ V}$$

From Table 1.1, $\phi_{A_1} = +0.6 \text{ V}$. Therefore, using (1.4.4),

$$\phi_{A_1, P_1} = 0.6 \text{ V} - (-0.29 \text{ V}) = 0.89 \text{ V}$$

Example 1.2. Calculate the contact potential ϕ_{P_2, N_2} of p -type silicon with $N_A = 10^{14} \text{ cm}^{-3}$ (denoted by P_2), to n -type silicon with $N_D = 10^{16} \text{ cm}^{-3}$ (denoted by N_2), at $T = 300 \text{ K}$.

From Table 1.1 and (1.4.2), $\phi_{P_2} = -0.234 \text{ V}$. From Table 1.1 and (1.4.3), $\phi_{N_2} = +0.353 \text{ V}$. Therefore, using (1.4.4),

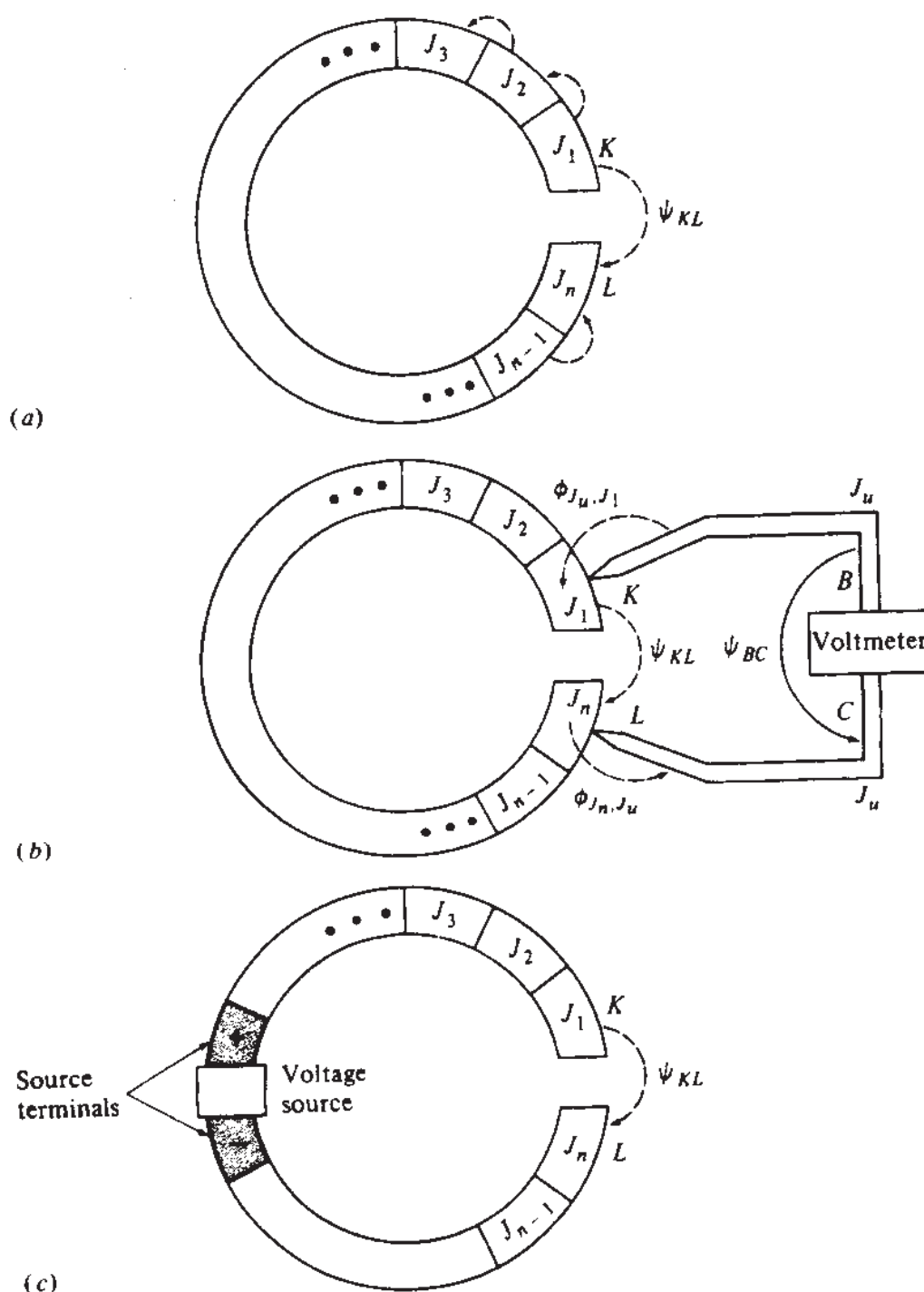
$$\phi_{P_2, N_2} = -0.234 \text{ V} - (+0.353 \text{ V}) = -0.587 \text{ V}$$

SEVERAL MATERIALS IN SERIES. We can now make an interesting observation concerning several materials in series, as in Fig. 1.14a; as before, all materials are assumed to be at the same temperature. If we express the potentials ψ_{KL} in terms of contact potentials in the loop, we have

$$\psi_{KL} = \phi_{J_1, J_2} + \phi_{J_2, J_3} + \cdots + \phi_{J_{n-1}, J_n} \quad (1.4.5)$$

and, using (1.4.4) repeatedly in the above equation, we get:

$$\psi_{KL} = (\phi_{J_1} - \phi_{J_2}) + (\phi_{J_2} - \phi_{J_3}) + \cdots + (\phi_{J_{n-1}} - \phi_{J_n}) \quad (1.4.6)$$

**FIGURE 1.14**

(a) Several materials in series; (b) the structure of (a) with a voltmeter attached; (c) the structure of (a) with a voltage source placed in series.

It is clear that, with the exception of ϕ_{J_1} and ϕ_{J_n} , each ϕ_{J_i} appears twice in the sum—once with a plus and once with a minus sign. Therefore

$$\psi_{KL} = \phi_{J_1} - \phi_{J_n} \quad (1.4.7)$$

A question arises: Can we measure potential ψ_{KL} with a common voltmeter? The answer is *no*, even if the voltmeter is assumed ideal (i.e., does not draw any cur-

rent from the circuit under measurement). This can be seen in Fig. 1.14*b*. Assume that the voltmeter leads are made by some material J_u . The potential difference ψ_{BC} between the two leads (which is what the voltmeter “sees”) is

$$\psi_{BC} = \phi_{J_u, J_1} + \psi_{KL} + \phi_{J_n, J_u} \quad (1.4.8)$$

Using (1.4.4) in the above equation,

$$\psi_{BC} = (\phi_{J_u} - \phi_{J_1}) + \psi_{KL} + (\phi_{J_n} - \phi_{J_u}) \quad (1.4.9)$$

and, using (1.4.7) in (1.4.9), we get

$$\psi_{BC} = 0 \quad (1.4.10)$$

In summary then, no matter how many materials are in the loop, the electrostatic potential difference between its two ends depends *only* on the first and the last material and cannot be measured by a common voltmeter.† For such measurements, special techniques must be used.^{14,15}

What now if a voltage source is inserted in the loop? This situation is illustrated in Fig. 1.14*c*. The source leads are assumed to be made of the same material. By going around the loop we have

$$\psi_{KL} = V_{\text{source}} + (\phi_{J_1} - \phi_{J_n}) \quad (1.4.11)$$

which should be compared with (1.4.7).‡

What would an ideal voltmeter like that of Fig. 1.14*b* measure if attached to points K and L in Fig. 1.14*c*? With ψ_{KL} as given by (1.4.11), (1.4.9) is still valid and we have

$$\psi_{BC} = V_{\text{source}} \quad (1.4.12)$$

which should be compared to (1.4.10).

It is now clear why contact potentials never seem to enter the picture when one works with circuits, provided all contacts are at the same temperature. As seen in the example leading to (1.4.12), such potentials cancel out, cannot be measured by common voltmeters, and do not enter into *circuit* equations. However, in investigating the physics of electronic devices, contact potentials must be taken into account; otherwise, unless a fortuitous cancellation takes place, neglecting contact potentials can lead to serious errors.

†The leads of a voltmeter are invariably made of the same material. The “academic” case where the leads are made of two different materials can be handled if one considers the various contact potentials within the voltmeter circuit (see Prob. 1.7).

‡Leads of voltage sources, like the leads of a voltmeter, are assumed to be made of the same material. The case of a source with leads made of different materials is easily handled by defining V_{source} as the voltage measured by our ideal voltmeter when attached to the terminals of the source. Then (1.4.11) and (1.4.12) will remain valid (see Prob. 1.8).

1.5 THE pn JUNCTION

The selective doping of silicon with n -type impurities in one region and with p -type impurities in an adjacent one produces what is known as a pn junction. Such a junction is shown in Fig. 1.15. In this figure, we have assumed that the doping changes abruptly from n -type to p -type at a boundary, and that the doping concentration is

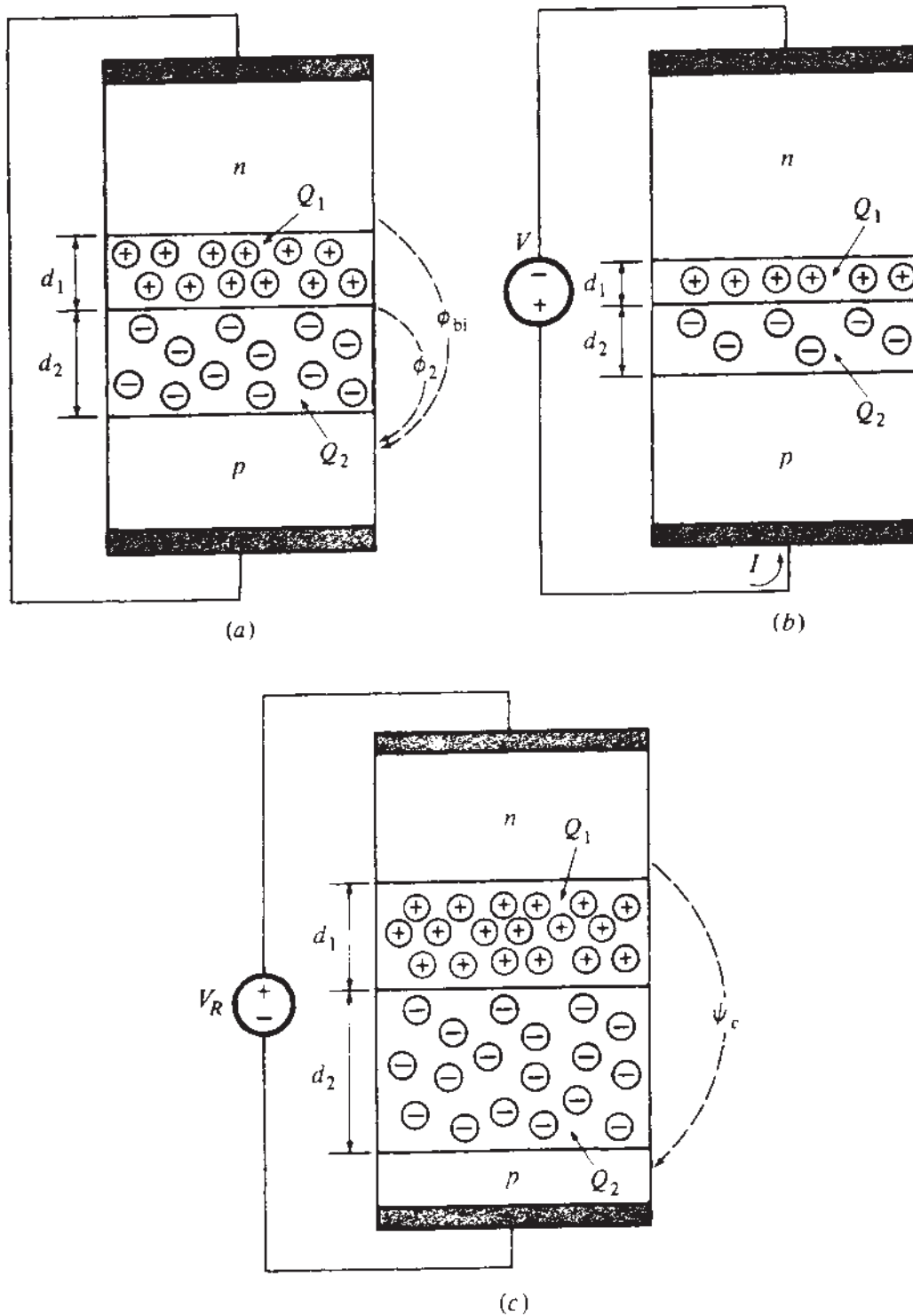


FIGURE 1.15

A pn junction with (a) zero bias, (b) forward bias, (c) reverse bias.

uniform on either side of the boundary; such a junction is referred to as an *abrupt*, or *step*, junction. In drawing Fig. 1.15, we have also assumed that the n region is more heavily doped than the p region. The two caps on the top and the bottom are made of a metal chosen so as to make “ohmic contacts” (contacts that conduct well in both directions†) with the semiconductor.

Let us first consider the case in Fig. 1.15a where the externally applied bias is zero. Free electrons are in high concentration in the n side and in low concentration in the p side. Thus, they tend to diffuse across the boundary between the two sides (downward in the figure), leaving behind positively charged immobile donor atoms as shown. Similarly, holes tend to diffuse from the p side, where they are in abundance, upward to the n side, leaving behind negatively charged immobile acceptor atoms. For simplicity it is often assumed that the above phenomena result in complete depletion of carriers over a region with sharply defined edges, as shown in the figure. This is called the *depletion approximation*, and the region is referred to as the *depletion region*. The more the uncovered donor atoms on the n side, the more their total charge; therefore, the stronger the electric field they produce, the stronger the attracting force they exert on electrons attempting to diffuse downward, and the stronger the repelling force on holes attempting to diffuse upward. Similarly, the more the uncovered acceptor atoms on the p side, the more they will attract holes which are attempting to diffuse upward, and the more they will repel electrons which are attempting to diffuse downward. Eventually in this process, enough impurity atoms are uncovered on both sides that the strong fields exerted by them inhibit a further net movement of carriers. The depletion region width is then fixed, and the current in the external wire is zero. It is interesting to note that the situation inside the structure will not change if the external wire is now cut. In fact, in this case it is easy to see from (1.4.6) and the associated discussion that the three contact potentials in the structure cancel each other out, and no electrostatic potential exists across the two open terminals.

In the absence of external bias, the potential across the depletion region (from deep in the n side to deep in the p side) will simply be the contact potential of the n -type material to the p -material. This potential is referred to as the *built-in potential* of the junction, and will be denoted by ϕ_{bi} . From Table 1.1 and (1.4.4), ϕ_{bi} is given by

$$\phi_{bi} = \phi_{Fp} - \phi_{Fn} \quad (1.5.1)$$

where ϕ_{Fp} and ϕ_{Fn} are the Fermi potentials of the p side and the n side, respectively. We will be interested in cases where one of the two regions is very heavily doped. We remind the reader that in such cases of a “degenerate” semiconductor, (1.4.2) or (1.4.3) cannot be applied. In practice, the value of ϕ_{bi} is sometimes chosen to provide best matching of expressions in which it appears to experimental results.

†How well current is conducted in either direction is something that cannot be deduced from the contact potential value. The theories for conduction in a metal-semiconductor contact are relatively involved.^{5,16}

Let the depths of the depletion region at the two sides of the junction in Fig. 1.15a be d_1 and d_2 , as shown. If the cross-sectional area of the junction (as seen from above) is A , the part of the depletion region in the n side has a volume equal to $(d_1 A)$, and therefore a total number of ionized atoms equal to $(d_1 A)N_D$. Since each of these atoms carries a charge $+q$, the total charge there, denoted by Q_1 , is

$$Q_1 = +q(d_1 A)N_D \quad (1.5.2)$$

Similarly, the total charge in the p side of the depletion region (which is due to ionized acceptor atoms) is

$$Q_2 = -q(d_2 A)N_A \quad (1.5.3)$$

For overall charge neutrality, we must have

$$Q_1 = -Q_2 \quad (1.5.4)$$

Let us now restrict our attention to the case of a "one-sided" step junction of the type n^+p , i.e., a step junction for which $N_D \gg N_A$; this last relation, along with (1.5.2) to (1.5.4) implies

$$d_1 \ll d_2 \quad (1.5.5)$$

which means that practically all the depletion region extends into the p side.

One can now apply Poisson's equation (1.2.13) in the depletion region, assuming all charge there is due to ionized impurity atoms (Appendix C). One finds that, in the case of $N_D \gg N_A$ considered here, practically all of the potential ϕ_{bi} across the junction will be dropped on the p side. Denoting by ϕ_2 the potential drop across the p side, we have:

$$\phi_2 \approx \phi_{bi} \quad (1.5.6)$$

Also, one finds that the distance d_2 , over which the acceptor atoms must be uncovered to support the potential drop $\phi_2 \approx \phi_{bi}$, is given by (Appendix C and Prob. 1.13):

$$d_2 = \sqrt{\frac{2\epsilon_s}{qN_A}} \sqrt{\phi_{bi}} \quad (1.5.7)$$

with ϵ_s given by (1.2.14).

Let us now define the charge *per unit area* on the p side, Q'_2 , as follows†:

$$Q'_2 = \frac{Q_2}{A} \quad (1.5.8)$$

†The results derived here will later be extended to cases where d_2 changes with horizontal position; in such cases, one must define Q'_2 at each horizontal position differentially as dQ'_2/dA .

From (1.5.3), (1.5.7), and (1.5.8) we obtain

$$Q'_2 = -\sqrt{2q\epsilon_s N_A} \sqrt{\phi_{bi}} \quad (1.5.9)$$

From (1.2.10) and (1.2.14), $\sqrt{2q\epsilon_s}$ has a value of $5.8 \times 10^{-16} \text{ C} \cdot \text{cm}^{-1/2} \cdot \text{V}^{-1/2}$ for silicon.

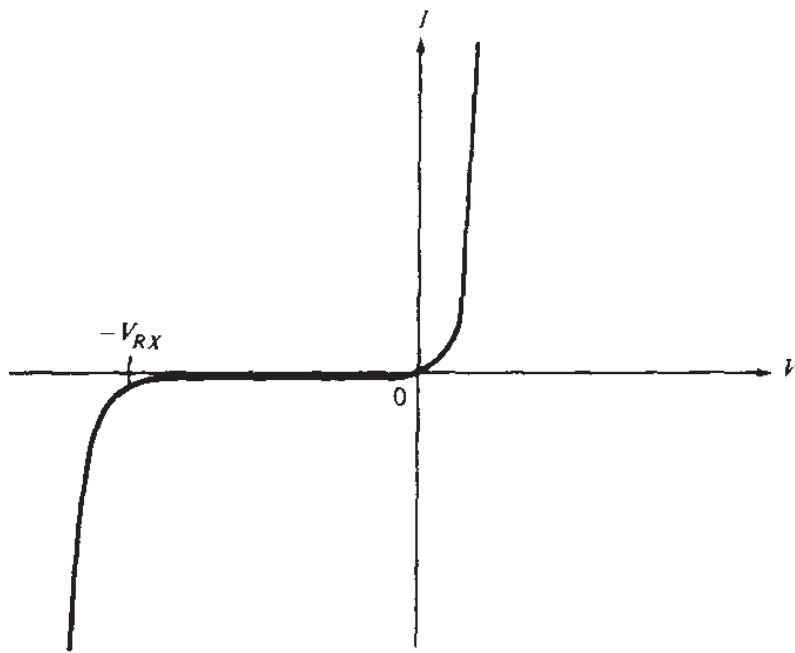
We now consider the application of a *forward* bias voltage $V > 0$, as shown in Fig. 1.15*b*. Equilibrium is now destroyed. The polarity of the voltage tends to produce a field opposite from that produced by the uncovered impurity atoms. Assuming that the contact potential of the *ohmic* contacts is not affected, the electrostatic potential across the depletion region will be reduced from ϕ_{bi} to $\phi_{bi} - V$, so that the sum of the electrostatic potentials around the loop remains zero. Thus, the total electric field is reduced at each point in the depletion region, and carriers find it easier to diffuse across the boundary. Electrons from the n side, being majority carriers there and thus in large supply, diffuse downward; holes from the p side, which are majority carriers there and therefore again in large supply, diffuse upward. The result of both of these movements is a positive current upward, as shown. For this to make sense, of course, one should recall that the assumption that *all* mobile carriers have left the depletion region, used previously, was just a convenient simplification; mobile carriers actually do exist in the “depletion” region and can support the flow of current. Assuming the current is not too large, and that the ohmic drop in the semiconductor outside the depletion region is negligible, it can be shown that the magnitude of the current will be given by^{2-10,12}

$$I = I_0 (e^{V/\phi_t} - 1) \quad (1.5.10)$$

where I_0 is a quantity dependent on junction geometry and physical parameters of the semiconductor material and is an increasing function of temperature, and ϕ_t is the thermal voltage given by (1.2.8).

The polarity of the source in Fig. 1.15*b* is such as to supply additional electrons to the n side, which can be viewed as “covering” some of the previously uncovered ionized donor atoms; hence the region where such ionized atoms are contained becomes narrower. From charge neutrality it is expected that some of the uncovered acceptor atoms on the other side will be covered as well. Hence the depletion region width is smaller than in Fig. 1.15*a*, which is consistent with our claim that the electric field at each point is less intense.

Consider now the case shown in Fig. 1.15*c*, where a *reverse* bias $V_R > 0$ is applied from the n -side terminal to the p -side terminal. Again, the structure will be in nonequilibrium. Now, the polarity of the applied voltage is such as to increase the electric field intensity at each point, compared to that in Fig. 1.15*a*. This polarity tends to move electrons upward away from the upper edge of the depletion region and holes downward away from the lower edge of that region. Thus the width of the depletion region will increase as shown. The larger the value of V_R , the wider the depletion region will become. If the n region is most heavily doped, overall charge neutrality dictates that such increases in width will take place mostly on the p side. From

**FIGURE 1.16**

Current vs. voltage for a *pn* junction.

this it can also be shown that the *p*-side depletion region will support most of the voltage drop.

We see that the direction of the field is now such that it no longer aids the flow of majority carriers from each side toward the junction boundary, as it did in the case of forward bias. Instead, some of the few holes (minority carriers) near the depletion region edge in the *n* side are swept downward by the field; similarly, some of the few electrons near the depletion region edge in the *p* side are swept by the field upward. This flow of minority carriers from each side causes a net external current in the opposite direction of that in Fig. 1.15*b*; and, since minority carriers are in short supply, this current is very small. Equation (1.5.10) remains valid for the reverse-bias case, with $V = -V_R$. For V_R larger than a few ϕ_p , the exponential becomes negligible and the current is seen to “saturate” at the value $-I_0$. For this reason, I_0 is often referred to as the *saturation current*. In practice, various “nonidealities” contribute to make the actual reverse-current magnitude larger than I_0 , and indeed somewhat dependent on V_R . The reverse-current magnitude is strongly dependent on temperature, doubling for every 10°C or even less of a temperature change. If V_R exceeds a certain value V_{RX} , the theory^{4-9,12} used to derive (1.5.10) fails. The junction conducts a large current from the *n* to the *p* side and is said to be in “reverse breakdown.” V_{RX} is called the *reverse-breakdown voltage*, and, for the junctions encountered on MOS integrated circuits, it is typically from 5 to 100 V, depending on fabrication details. A complete *I-V* characteristic of a *pn* junction, including breakdown, is shown in Fig. 1.16. Most *pn* junctions on MOS integrated circuits are kept reverse-biased with

$V_R < V_{RX}$. Their reverse current can often be neglected except in certain critical applications,[†] or unless the temperature is very high. Note, however, that the reverse current in Fig. 1.15c is a manifestation of nonequilibrium. This current, no matter how small, provides communication between the junction and the external battery, which drastically alters the conditions that were present at equilibrium under zero bias. Thus, for example, relations that were valid in equilibrium, such as (1.2.6) or (1.2.7), will not be valid under nonzero bias in general.

Under the reverse bias V_R , the electrostatic potential across the depletion region (ψ_c in Fig. 1.15c) will increase by V_R , so that the potential drops around the loop add up to zero again (the contact potentials at the ohmic contacts are considered unchanged). Thus

$$\boxed{\psi_c = \phi_{bi} + V_R} \quad (1.5.11)$$

The new depletion region width will be given by (1.5.7) if the new potential ψ_c is used instead of ϕ_{bi} (Appendix C):

$$d_2 = \sqrt{\frac{2\epsilon_s}{qN_A}} \sqrt{\psi_c} \quad (1.5.12)$$

Similarly, the new depletion region charge per unit area on the p side will be given by the following equation instead of (1.5.9):

$$\boxed{Q'_2 = -\sqrt{2q\epsilon_s N_A} \sqrt{\psi_c}} \quad (1.5.13)$$

We now consider the concept of the small-signal capacitance of the reverse-biased pn junction. If the reverse bias is increased by a small amount ΔV_R , the depletion region width must increase on both sides. The charge that must be moved for this to happen must flow through the external circuit. Specifically, additional donor atoms will be uncovered in the n side, increasing the total charge there. For this to happen, electrons must be removed by the external circuit from near the depletion region's top edge. Thus, a negative charge, say $-\Delta Q$, leaves the junction from the top terminal. This can equivalently be described by saying that an opposite *positive* charge $+\Delta Q$ flows *into* the junction through the top terminal. Similarly, since the reverse bias has increased, additional acceptor atoms must be uncovered in the p side. For this to happen, the external circuit must cause positively charged holes to be removed from near the depletion region's bottom edge. This implies a positive charge $+\Delta Q$ flowing out

[†]For example, the "standby" current of circuits can be determined by leakage and can limit battery life in portable applications.

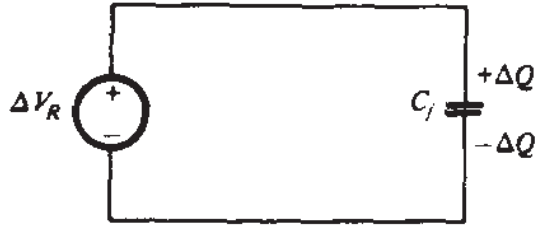


FIGURE 1.17

Small-signal equivalent circuit of a reverse-biased *pn* junction driven by a voltage source.

of the bottom terminal or, equivalently, a *negative* charge $-\Delta Q$ flowing *into* the bottom terminal. Thus the changes of the charges Q_1 and Q_2 in Fig. 1.15c will be

$$\Delta Q_1 = +\Delta Q \quad (1.5.14)$$

$$\Delta Q_2 = -\Delta Q \quad (1.5.15)$$

Assume now that a voltage source of value equal to the *change* of the reverse bias ΔV_R is connected across a capacitor as in Fig. 1.17, and that its capacitance value is chosen so that the charges on the two plates are $+\Delta Q$ and $-\Delta Q$ (i.e., they are equal to the incremental charges that entered the junction through the top and the bottom terminal, respectively, when the reverse bias was changed from V_R to $V_R + \Delta V_R$). Then this capacitance represents the *small-signal capacitance* of the reverse-biased *pn* junction. If this capacitance is denoted by C_j , we have

$$C_j = \frac{\Delta Q}{\Delta V_R} \quad (1.5.16)$$

Using (1.5.15), and refining the definition of C_j by letting the finite differences become differentials, we have

$$C_j = - \frac{dQ_2}{dV_R} \quad (1.5.17)$$

Finally, dividing both sides by the cross-sectional area A , we obtain the *small-signal capacitance per unit area* C_j/A , denoted by C'_j :

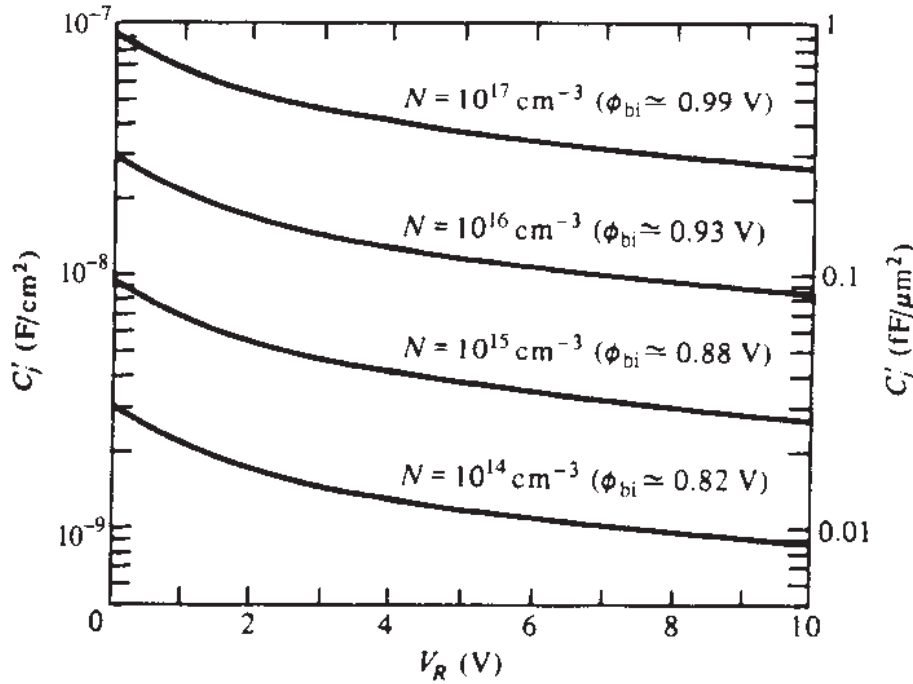
$$C'_j = - \frac{dQ'_2}{dV_R} \quad (1.5.18)$$

Using now (1.5.11) and (1.5.13) in the above gives

$$C'_j = \frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{V_R + \phi_{bi}}} \quad (1.5.19)$$

which, using (1.5.11) and (1.5.12), can also be written as

$$C'_j = \frac{\epsilon_s}{d_2} \quad (1.5.20)$$

**FIGURE 1.18**

One-sided step-junction small-signal capacitance per unit area vs. reverse-bias voltage, with doping concentration of the lightly doped side as a parameter. The heavily doped side is assumed degenerate; corresponding approximate values of built-in voltage are indicated.⁵ $T = 300$ K.

This equation is plotted in Fig. 1.18, where approximate values for ϕ_{bi} are given⁵ assuming the heavily doped side is degenerate. If C'_{j0} denotes the value of C'_j for $V_R = 0$, (1.5.19) can be easily converted to the following:

$$C'_j = \frac{C'_{j0}}{\sqrt{(V_R/\phi_{bi}) + 1}} \quad (1.5.21)$$

Again it is emphasized that the above results have been derived for a one-sided, abrupt junction. For more complex doping transitions from the n to the p side, it can be shown⁵ that C'_j can be approximated by:

$$C'_j = \frac{C'_{j0}}{[(V_R/\phi_{bi}) + 1]^{\alpha_j}} \quad (1.5.22)$$

where the expression for C'_{j0} will depend on the details of the transition from n type to p type (these details form the so-called *profile* of the junction). The value of α_j is $1/2$ for the one-sided abrupt junction we have discussed, and $1/3$ for the linearly graded junction (one in which the net impurity concentration varies linearly with vertical position, being zero at the boundary between the two sides). For practical profiles (which are neither exactly abrupt nor exactly linearly graded) one often chooses the values of the parameters α_j , C'_{j0} , and ϕ_{bi} so as to obtain best fit of (1.5.22) to measurements. The values obtained this way are said to be “extracted” from measurements.

1.6 OVERVIEW OF THE MOS TRANSISTOR

1.6.1 Basic Structure

The basic idea on which the MOS transistor is based is over half a century old and is due to J. E. Lilienfeld, who obtained the first patents in the early 1930s¹⁷; see Fig. 1.19. Other patents were obtained by O. Heil.¹⁸ Laboratory studies were performed in the late forties (see, for example, Ref. 19) but the device remained in the laboratory for over a decade following that. Then in the early sixties the MOS transistor “took off” following the demonstration of working devices by Kahng and Atalla,²⁰ the development of techniques for reliably growing oxides,²¹ and the establishment of basic theories of operation.^{22–25} A historical review of early work, including the development of practical fabrication techniques, can be found elsewhere.²⁶ Modern fabrication techniques are described in a number of books.^{27–29}

We now offer a simple preview³⁰ of the MOS transistor, which will help introduce the detailed material in the rest of this book.

A simplified structure of an *n*-channel MOS transistor is shown in Fig. 1.20 (the names *n*-channel and MOS will be discussed shortly). The transistor is formed on a *p*-type silicon *body* (or *substrate*). Typical doping concentrations for the body are 10^{16} to 10^{18} cm⁻³. The dopant concentration will be assumed uniform throughout the body, until further notice.

The center part of the structure is covered by an insulator (typically silicon dioxide, which is often referred to simply as *oxide*), usually of 35 to 100 Å thickness. The body interface to the oxide is often called the *surface*.

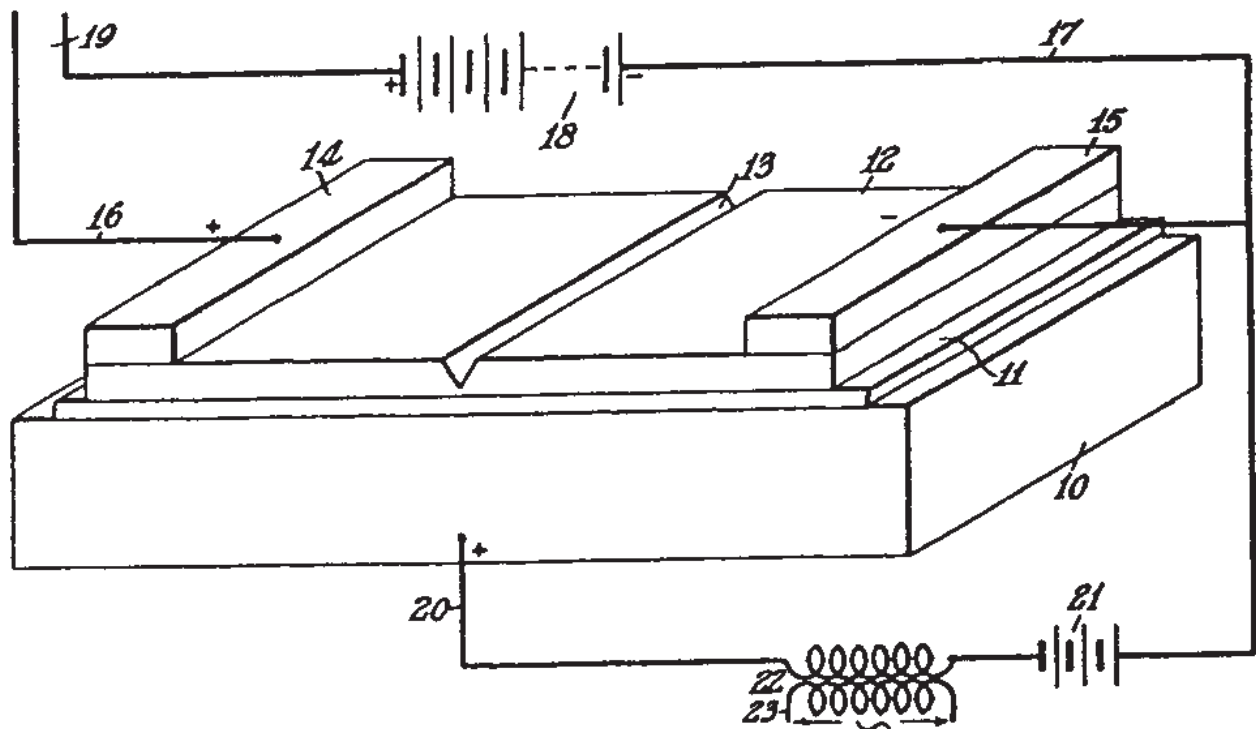


FIGURE 1.19

A figure from J. E. Lilienfeld's U.S. patent 1,900,018 (filed March 28, 1928; granted 1933). The structure should be turned upside down to correspond to the orientation used in this book (e.g., Fig. 1.20).

A low-resistivity electrode, called the *gate*, is formed on top of the oxide. Contemporary processes commonly use polycrystalline silicon (*polysilicon*, or *poly*, for short) for the gate. This material, although silicon, is not a single crystal. Rather, it consists of many regions within each of which there is a regular array of atoms, and this regularity is broken at the boundaries between adjacent regions. The polysilicon material is heavily doped *p* or *n* type (e.g., 10^{20} cm^{-3}). The two regions shown on the sides are formed by implanting donor atoms, with the gate acting as a mask against the implant; this mask receives the donor atoms itself and prevents them from landing under it. Thus the gate is heavily doped and exhibits low resistivity. Donor atoms land in the substrate just outside the “shadow” of the gate, and form the two n^+ (heavily doped *n*) regions indicated as *source* and *drain* in Fig. 1.20; these regions are typically 0.04 to 0.2 μm deep. The heavy doping results in low resistivity for these regions, since the abundance of free electrons in them is available for conduction. Subsequent high-temperature fabrication steps cause a diffusion of the dopant atoms both vertically and laterally. This *lateral diffusion* causes the source and drain regions to extend slightly under the gate as shown in the figure. The resulting overlap distance is typically 0.02 to 0.1 μm .

The region between the source and drain is called the *channel*. The channel *width* W and *length* L of individual transistors can vary greatly (from a fraction of a micrometer to several hundred micrometers), depending on circuit design needs. In digital circuits, L is normally kept at the minimum value possible.

As we will see below, if the gate potential is made sufficiently positive with respect to other parts of the structure, electrons can be attracted directly below the insulator (near the “surface” of the body). These electrons can come through the n^+ regions, where they exist in abundance, and can fill the channel between them; for this reason the device in Fig. 1.20 is referred to as an *n-channel* device (the opposite-type device, called *p-channel*, has holes in its channel and will be considered later). The number of electrons in the channel can be varied through the gate potential. This can cause a variation of the “strength” of the connection between the two n^+ regions, resulting in transistor action. If the two n^+ regions are biased at different potentials, the lower-potential n^+ region acts as a source for electrons, which then flow through the

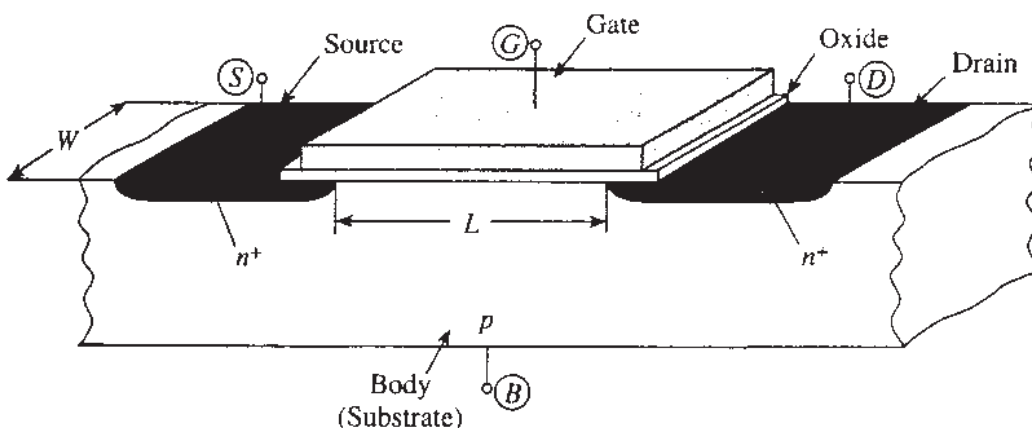


FIGURE 1.20

Simplified structure of an *n*-channel MOS transistor.

channel and are drained by the higher-potential n^+ region. It is thus common to call the lower-potential n^+ region *source*, and the higher-potential one *drain*.

The source, gate, drain, and body regions can be contacted through terminals attached to them, as shown schematically in Fig. 1.20. We will denote these terminals by S , G , D , and B , respectively.

The first successful MOS transistor used metal for the gate material and silicon dioxide for the insulator. It is for this reason that the device was named *MOS* transistor, with MOS standing for *metal oxide semiconductor*. Other acronyms are *MOST* (for MOS Transistor), *MOSFET* (for MOS Field-Effect Transistor), and *IGFET* (Insulated-Gate Field-Effect Transistor). The latter name originated to distinguish the device from the junction-gate field effect transistor, in which the gate is separated from the rest of the structure by a pn junction. Among the three acronyms, IGFET is the most general, since it does not specify the material used for the gate or the insulator. This acronym, though, is not in wide use. Today, the popular acronyms MOST and MOSFET have come to mean the same as IGFET and do not imply that metal and silicon dioxide are necessarily used for the gate and the insulator.

Often n -channel and p -channel MOS transistors are referred to simply as *nMOS* and *pMOS* transistors, respectively. Both types of devices are available on chips made using CMOS (complementary MOS) technology. Partial views of three types of CMOS chips are shown in Fig. 1.21. In Figure 1.21a, we show the result of the so-called “local oxidation of silicon” (LOCOS) CMOS process.^{27–31} Characteristic of this process is the gradual transition of the oxide from thin (under the gates) to thick, resulting in the so-called “bird’s beak” shape. The thick oxide is needed so that wiring running above it (not shown) cannot accidentally create a parasitic channel underneath it. Heavily doped “channel stop” regions are used under the thick oxide for the same purpose. In the CMOS process version shown in Fig. 1.21a, the p substrate is common to all n MOS devices on the chip and serves as an isolation region between them. The pn junctions formed between the p substrate and the n sources and drains are normally kept reverse-biased, by connecting the substrate to the most negative potential in the circuit. The p MOS devices are contained within n -type wells as shown; for this reason, this version of the CMOS fabrication process is referred to as an *n-well process*. To keep the junctions formed between the n well and the p sources and drains reverse-biased, the well must be connected to an appropriate potential which, for digital circuits, is the most positive potential available in the circuit. Contacts to the source and drain regions are made through oxide cuts called *contact windows* (not shown). Contacts are also made to the substrate and well regions. An oxide layer (not shown), called *passivation layer*, is deposited on top of the devices to protect them from their environment.

The minimum possible feature dimensions are determined by the capabilities of fabrication equipment and reliability considerations. Extra care is taken to avoid a condition known as “latchup,”³¹ in which some of the diodes formed by pn junctions may become accidentally forward-biased, thereby initiating bipolar transistor action and activating a type of positive feedback involving parasitic $pnpn$ structures in Fig. 1.21a. If latchup occurs, such diodes can be locked into an on condition,

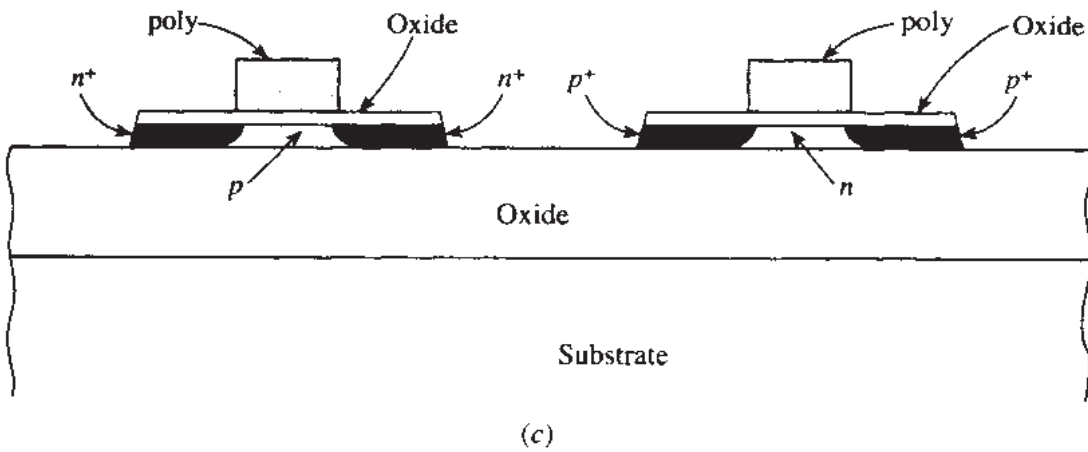
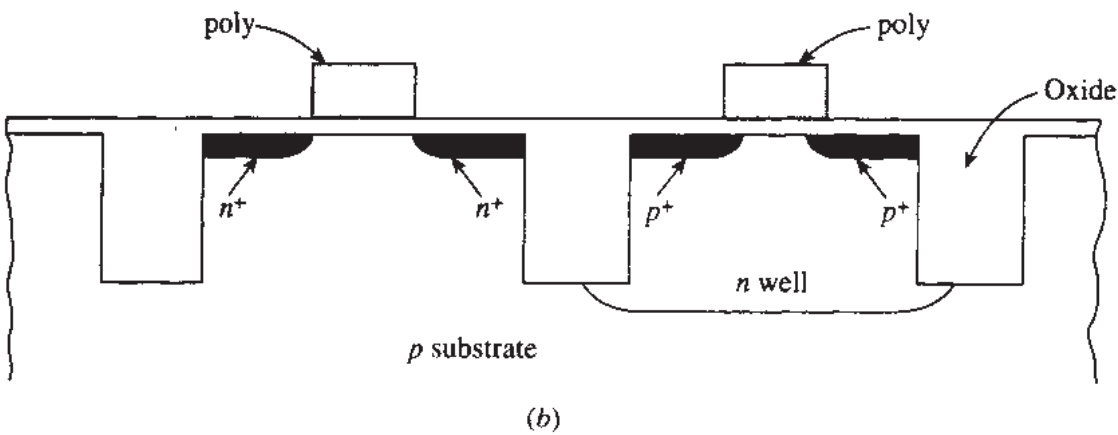
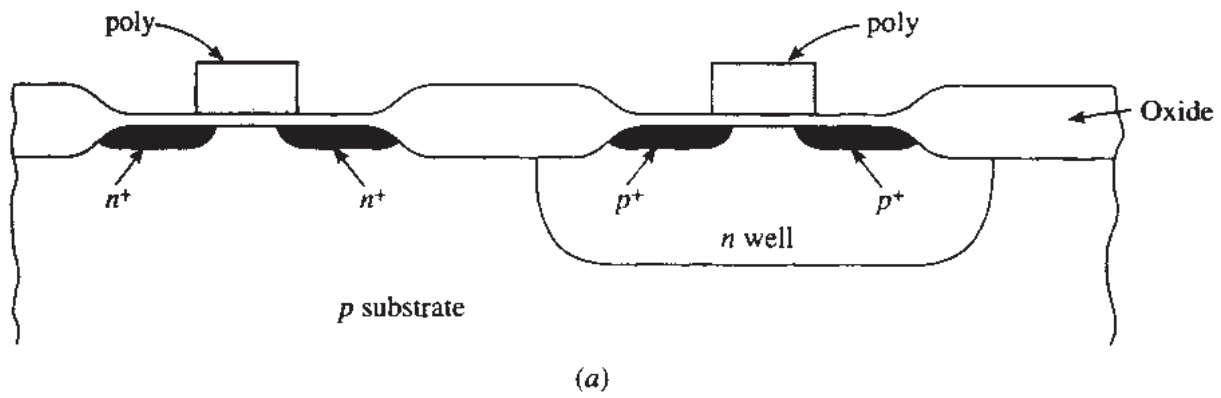


FIGURE 1.21
CMOS processes (simplified). (a) LOCOS, (b) STI, (c) SOI. Contacts to the various regions are not shown.

and excessive currents can flow which can interfere with proper operation and can destroy the chip. A sufficient distance between the various regions must be maintained to prevent latchup. It is also desirable to keep the substrate resistivity low, in order to prevent the formation of significant voltage drops due to parasitic currents in it, since such voltage drops can give rise to, and can sustain, latchup. However, this requires a high substrate doping concentration, which can have an undesirable effect on MOS transistor operation, as we will see. This conflict is resolved in some processes by using a shallow, lightly doped p-type “epi” region (so called because it is formed by a process called *epitaxy*^{27–29}), on top of a heavily-doped (p^+) substrate. Thus, the body of the MOS devices is made of lightly doped material, whereas the material underneath it is heavily doped to keep the parasitic resistances (and also the current gains of the parasitic bipolar devices) low.

Figure 1.21*b* shows a different type of CMOS process, known as *shallow-trench isolation* (STI).^{27–29} This is a more advanced process which is becoming prevalent. As shown, a “trench” filled with oxide isolates devices from one another, thereby allowing the devices to be packed closer to each other without compromising latchup immunity. Even more complete isolation is possible in silicon on insulator (SOI) processes,³² in which each device is surrounded by oxide and is thus completely isolated from its neighbors, as shown in Fig. 1.21*c*. In SOI processes the substrate of each device is normally left floating, and this requires special modeling considerations. In this book, we do not cover the modeling of floating-substrate SOI devices; the interested reader is referred to the literature.³² Also, we do not discuss devices made using special processes for high-voltage and power applications.³³

The actual values of channel width and length after fabrication (shown as W and L in Fig. 1.20), can be different from those specified during the chip “layout” process, which takes place on a computer screen during chip design. Several reasons contribute to this difference. For example, during high-temperature fabrication steps the source and drain regions diffuse not only downward in Fig. 1.20, but also laterally, thus changing the initial distance between them. The bird’s beak effect in LOCOS processes, which also exists along the width direction (not shown in Fig. 1.21*a*), is one reason that the channel width also becomes different from that specified during layout. In general, we have:

$$W = W_m - \Delta W \quad (1.6.1)$$

$$L = L_m - \Delta L \quad (1.6.2)$$

where W and L are the *actual* channel width and length, respectively (sometimes called *effective*, or *electrical*, width and length), W_m and L_m are the corresponding layout values (often referred to as *mask* or *drawn* values), and ΔW and ΔL are the corrections that must be applied to arrive at the true dimensions. These corrections are normally positive, and can be up to several tenths of a micrometer each. In this book, the terms *channel width* and *channel length* will exclusively refer to *actual* values of these quantities (i.e., W and L), unless specifically mentioned otherwise.

As already mentioned, the source-body and drain-body junctions are normally kept reverse-biased. This guarantees that only a very small reverse-bias “leakage”

current flows in the pn junctions (typically less than 1 pA for minimum-size regions at room temperature; note, though, that this current increases with temperature, roughly doubling every 8 to 10°C). The current flowing through the gate insulator is even smaller than the above current, in fact by orders of magnitude. Thus, unless stated otherwise, we will be assuming that leakage currents are negligible. The voltages between any two terminals of the transistor will be assumed to be sufficiently low to prevent breakdown, which otherwise can occur in reverse-biased junctions (Sec. 1.5), between the channel and the body, or in the oxide.

One last assumption that will hold for all the discussions in this book is that all parts of a transistor operate at the same temperature. Unless otherwise mentioned, this will be assumed to be “room temperature,” taken as 300 K. The effects of temperature on transistor behavior at other temperatures will be considered separately.

1.6.2 A Qualitative Description of MOS Transistor Operation

An n -channel MOS transistor with external voltages applied is shown in Fig. 1.22. The condition shown is for a sufficiently positive gate-source voltage V_{GS} , so that positive charges are placed on the gate. These repel the holes from the surface, leaving the latter depleted of holes. The resulting depletion region contains a number of negatively charged acceptor ions, a situation similar to that on the p side of the pn junction in Fig. 1.15. Further, V_{GS} is assumed to be sufficiently positive to even make the surface attractive to electrons; the latter can easily enter through one or both of the two n^+ regions, where they are in large supply. This situation is called *inversion*, since the surface of the p -type body, which normally would have a large

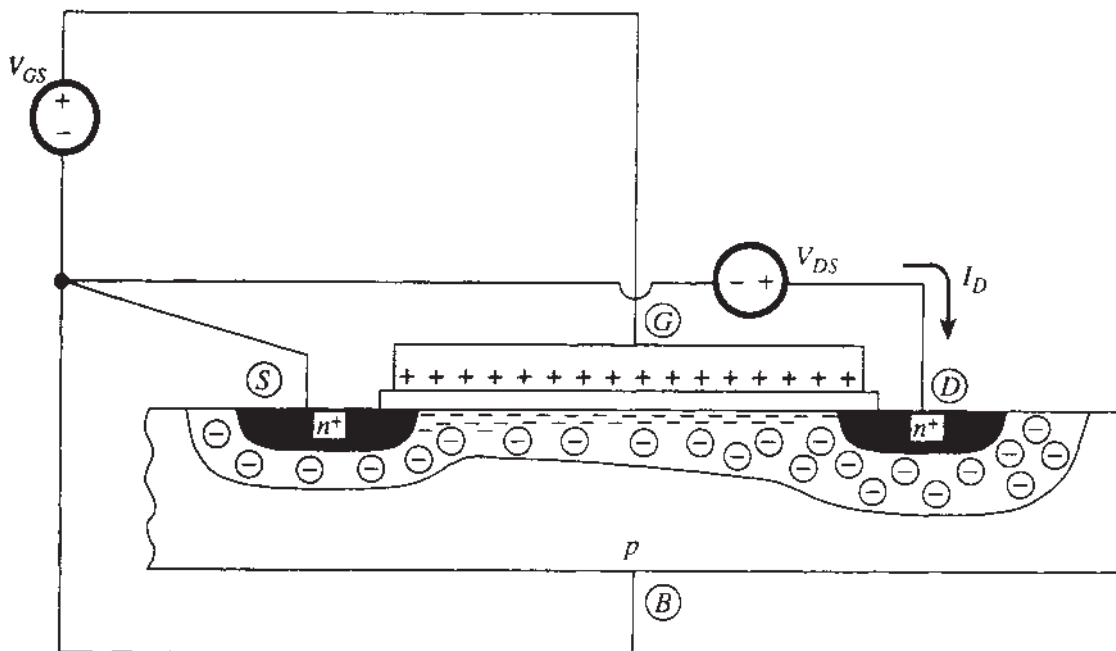


FIGURE 1.22
An n -channel MOS transistor under bias in the inversion region.

concentration of holes, now has plenty of electrons. The layer of electrons at the surface is called an *inversion layer*.

The source and the drain form two *np* junctions with the body. As we have assumed above, these are reverse-biased. As has been shown in Fig. 1.15, in a *pn* junction the resulting depletion region extends to both the *n* and *p* sides. However, in the *n*MOS transistor of Fig. 1.22 the part inside the *n* regions is much shallower, since the doping there is much heavier (see Sec. 1.5); thus, that part is not shown for simplicity. In drawing Fig. 1.22, we have assumed that the drain potential is more positive than the source potential. Thus, the reverse bias across the drain-body *np* junction is larger, and the depletion region shown for that junction is deeper. As a result, there are larger numbers of negatively charged acceptor atoms around the drain than there are near the source. This means that fewer electrons are needed in the channel near the drain to balance the positive charges on the gate. It is for this reason that the concentration of electrons is shown to decrease as the drain is approached in Fig. 1.22. The largest electron concentration is found near the source. The larger the value of the gate potential, the more the electrons and the “heavier” the inversion is at that point. A few volts of variation of the gate potential can vary the population of electrons there by several orders of magnitude. Although such variation is continuous, we often say that as the gate potential is raised, we go from *weak* inversion to *moderate* inversion, and eventually to *strong* inversion. It will be seen that, when appropriately done, this division into three regions is convenient, as distinct types of behavior are observed in each of the three regions.

The potential difference V_{DS} between the drain and the source is positive and appears across the inversion layer. It causes electron movement: electrons enter from the source, pass through the channel, and are drained by the drain. The electrons start relatively slowly at the source, and as they approach the drain they speed up; in this way, constant current can be maintained along the channel, although the electron concentration varies along the channel's length. The movement of negative charges toward the right corresponds to positive current I_D from the drain, through the channel, to the source, as shown by the arrow in Fig. 1.22.

Assume now that V_{DS} starts at zero and is gradually increased. The resulting drain current will also increase, as shown qualitatively in Fig. 1.23. For small values of the drain potential, the effect of the drain potential on the drain current is large; for large enough values of the drain potential, though, the current gradually tends to saturate as indicated by the part of the curve toward the right in Fig. 1.23. This happens when the drain potential is so large that it drains all electrons that can be supplied by the channel, for a given gate potential. Two regions can thus be distinguished, marked *nonsaturation* and *saturation* in the figure. Although the transition between nonsaturation and saturation is smooth, their boundary is sometimes taken to be at a specific point for convenience.

In Fig. 1.22, a short is shown between the body and source terminals. Assume now that this connection is broken, and that a voltage source is inserted in such a way as to make the body potential negative with respect to the source, as shown in Fig. 1.24. This will cause the population of electrons in the channel to decrease, which will be manifested externally as a decrease in I_D . This phenomenon is referred to as *body effect*. It can be understood qualitatively as follows: The body is a conductive struc-

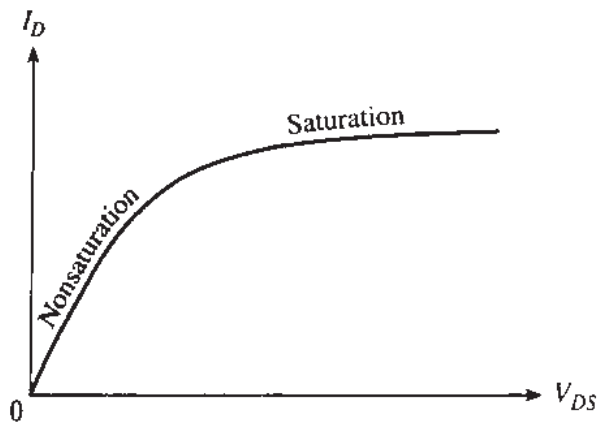


FIGURE 1.23

Typical behavior of drain current vs. drain-source voltage, for a fixed gate-source voltage.

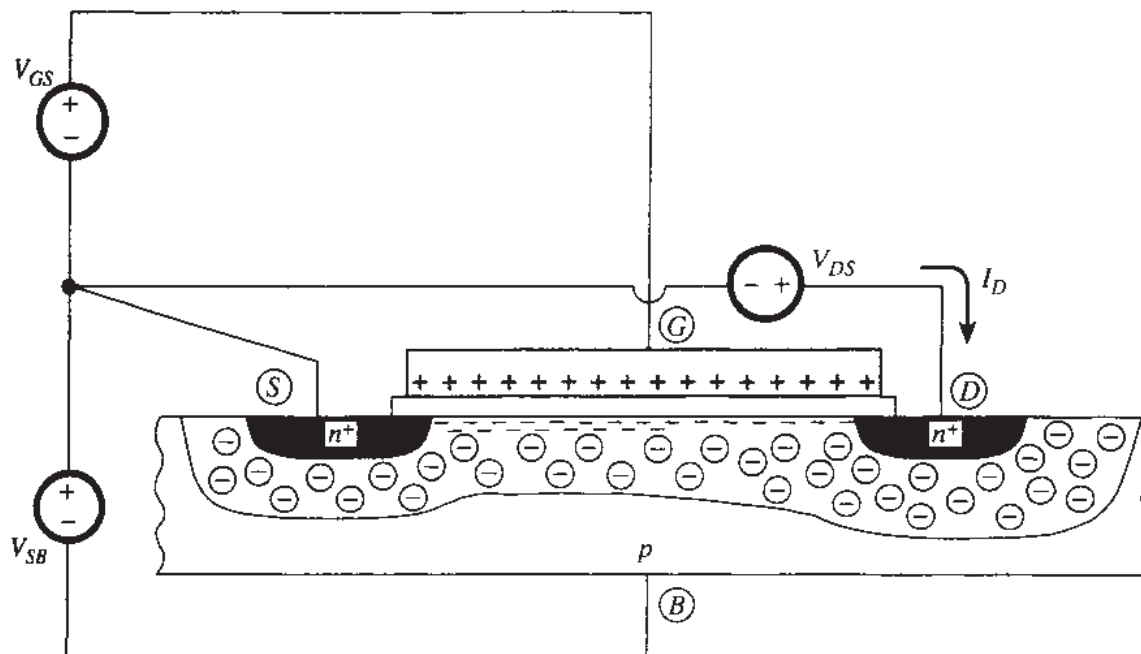


FIGURE 1.24

The connection of Fig. 1.22, modified to insert a negative bias between body and source.

ture, separated from the channel by an “insulating” region (the depletion region), just as the gate is separated from the channel by the oxide. The body, then, qualitatively acts as another “gate” (in fact, it is sometimes referred to as a *back gate*): applying a negative potential to it reduces the population of electrons, just as would be the case with the application of a negative potential on the gate. A more rigorous explanation of the body effect must await our detailed discussions in subsequent chapters.

1.6.3 A Fluid Dynamical Analog

To help increase intuition about the MOS transistor, we now present a fluid dynamical analog.^{34,35} In our analog, electrons correspond to water molecules. Electric current corresponds to net flow of water. The source and drain correspond to two large tanks, filled with water up to a certain level each. As water is moved from one tank to

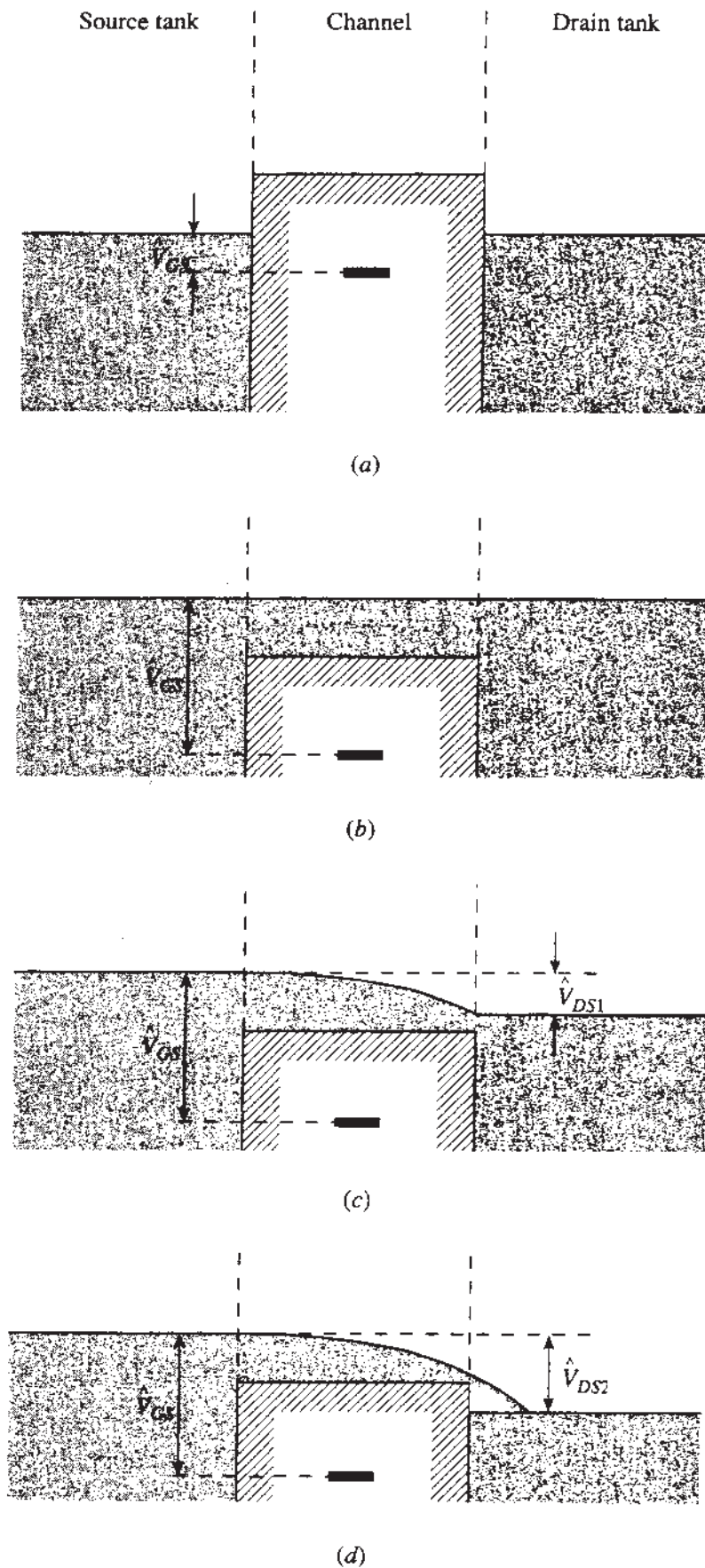
the other, the level of water in each tank is maintained constant through external means. This corresponds to the potentials at the source and drain of a transistor being held constant despite current flow. We will assume that the level of water in the "source" tank is fixed throughout our discussion. The two tanks are separated by a piston corresponding to the gate, as shown in Fig. 1.25a. A "handle," shown by a thick short line at the center, is attached to the piston. The depth at which this handle is with respect to the "source" level will be denoted by \hat{V}_{GS} and corresponds to the gate-source voltage V_{GS} . In this water analog, quantities corresponding to potentials increase *downward* in the figure. In particular, the lower the handle (and thus the piston), the larger the value of \hat{V}_{GS} . Similarly, the level of water in the "drain" tank with respect to the source level will be denoted by \hat{V}_{DS} and corresponds to the drain-source voltage. The lower the "drain" level, the larger the value of \hat{V}_{DS} .

In Fig. 1.25a, \hat{V}_{GS} is very low. The "channel" is cut off, and no communication exists between source and drain. Now assume \hat{V}_{GS} is increased considerably (i.e., the piston is moved *downward*), as shown in Fig. 1.25b. The channel is now filled with the water. Communication between the source and drain is now possible, but no flow is observed in the steady state, since $\hat{V}_{DS} = 0$. Now if \hat{V}_{DS} is increased as in Fig. 1.25c, flow is observed as shown. The flow increases as \hat{V}_{DS} is increased further, until "saturation" is reached; it is easy to see that in saturation (Fig. 1.25d), further increases in \hat{V}_{DS} do not affect the flow. The water enters slowly from the source and, as it approaches the drain, moves faster, in order to maintain a fixed flow despite the fact that the amount of water is less near the drain. Again, an analogy exists here to the simplified picture of transistor operation in saturation.

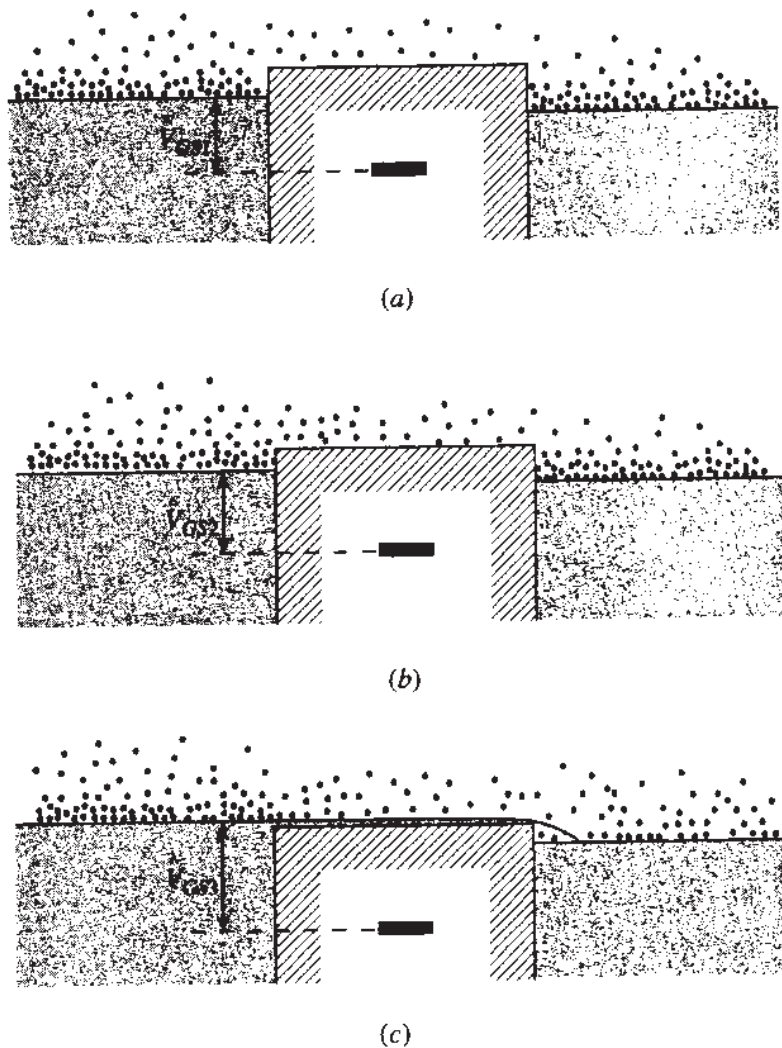
It is evident from the figure that, for a given \hat{V}_{DS} , the flow will increase if \hat{V}_{GS} is increased, and will decrease if \hat{V}_{GS} is decreased. If \hat{V}_{GS} is decreased to the point that the top surface of the piston is at, or above, the source surface level, direct flow of the fluid is prevented. Water molecules can still, though, flow from left to right. This can be seen if one takes into account the diffusion of water vapors,³⁵ as illustrated in Fig. 1.26a. This situation corresponds to weak inversion. As shown, the vapor concentration is maximum at the water surface and decreases as one moves vertically away from it (in fact, it can be shown that it decreases exponentially). Thus, if the water surface in the drain tank is lower, as shown in Fig. 1.26a, at any given horizontal plane above the piston the vapor concentration will be decreasing as one goes from left to right over the channel. Vapor thus diffuses from left to right, carrying a minute "current," even if the position of the middle shaft is slightly above the source water surface. However, a very small drop in the position of the middle shaft (i.e., an increase in \hat{V}_{GS}) can drastically increase the "current," since the vapor concentration over the source, at points slightly above the level of the shaft, will increase exponentially (Fig. 1.26b). One thus might expect an exponential dependence of the "current" on \hat{V}_{GS} .

For large \hat{V}_{DS} (low drain water level), the concentration on the right at the level of the shaft's surface becomes negligible, and the "current" assumes a value largely independent of \hat{V}_{DS} , reaching saturation.

If the top surface of the piston is only *very* slightly *below* the source level, both the fluid and its vapors can contribute significantly to the flow (Fig. 1.26c). Notice that vapors would also exist in the case of very positive \hat{V}_{GS} shown in Fig. 1.25. How-

**FIGURE 1.25**

Fluid dynamical analog of MOS transistor operation in cutoff and strong inversion.³⁴ (a) Cutoff; (b) with channel strongly inverted, but no current since $V_{DS} = 0$; (c) nonsaturation with nonzero V_{DS} ; (d) saturation. (Figure reprinted with permission from Y. Tsividis, *Mixed Analog-Digital VLSI Devices and Technology*, McGraw-Hill, 1996.)

**FIGURE 1.26**

Fluid dynamical analog of MOS transistor operation in (a and b) weak inversion and (c) moderate inversion.³⁵ (Figure reprinted with permission from Y. Tsividis, *Mixed Analog-Digital VLSI Devices and Technology*, McGraw-Hill, 1996.)

ever, they were not considered there because their contribution was negligible in comparison to the much larger flow of liquid water.

1.6.4 MOS Transistor Characteristics

Let us finally look at a set of typical transistor characteristics. In Fig. 1.27a we show plots of I_D (on a *logarithmic* axis) versus V_{DS} , with V_{GS} as a parameter, and with the body shorted to the source, for a specific transistor. The logarithmic axis is used to reveal the several orders of magnitude of I_D , over which control is possible through V_{GS} . The regions of inversion have been roughly marked in terms of V_{GS} (the limits between the regions will be discussed later in this book). As V_{GS} or V_{DS} is reduced, the current in the channel can eventually become so small that it is masked by the leakage current of the reverse-biased drain-body junction (or even the leakage from the inversion layer to the substrate). This can be seen near the bottom of the plot. The

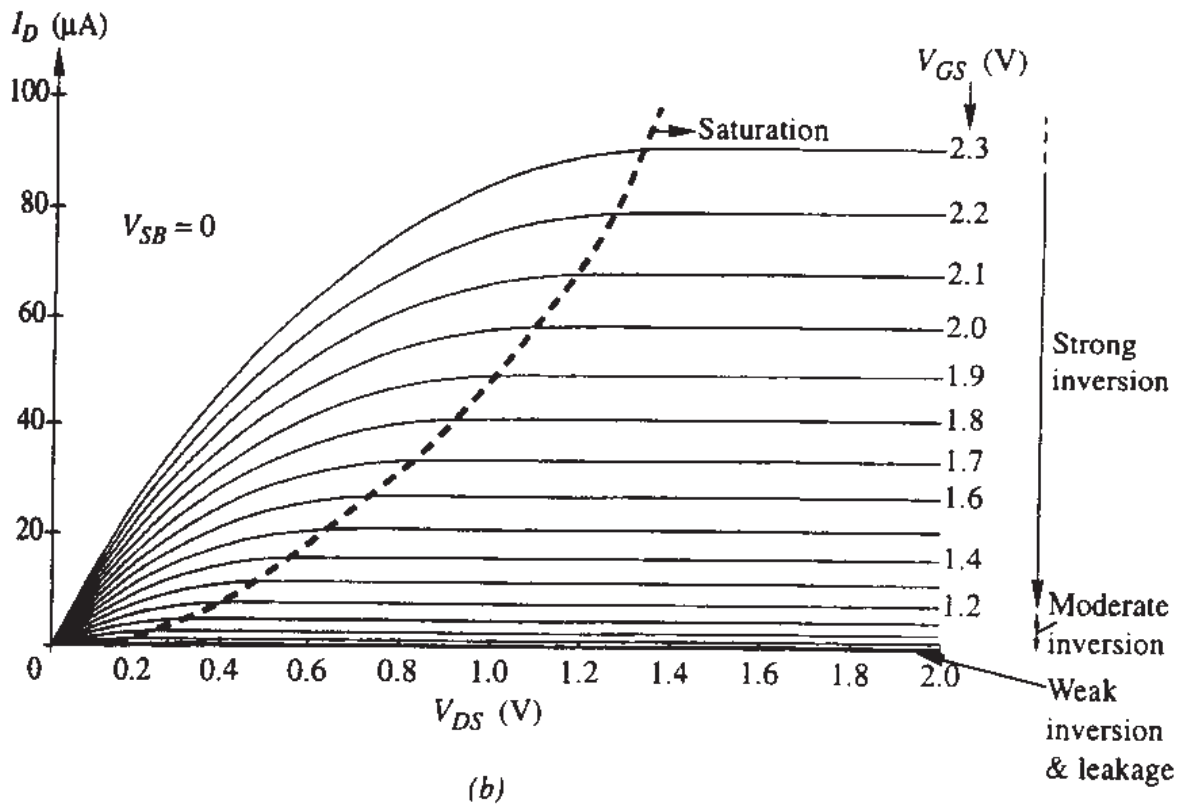
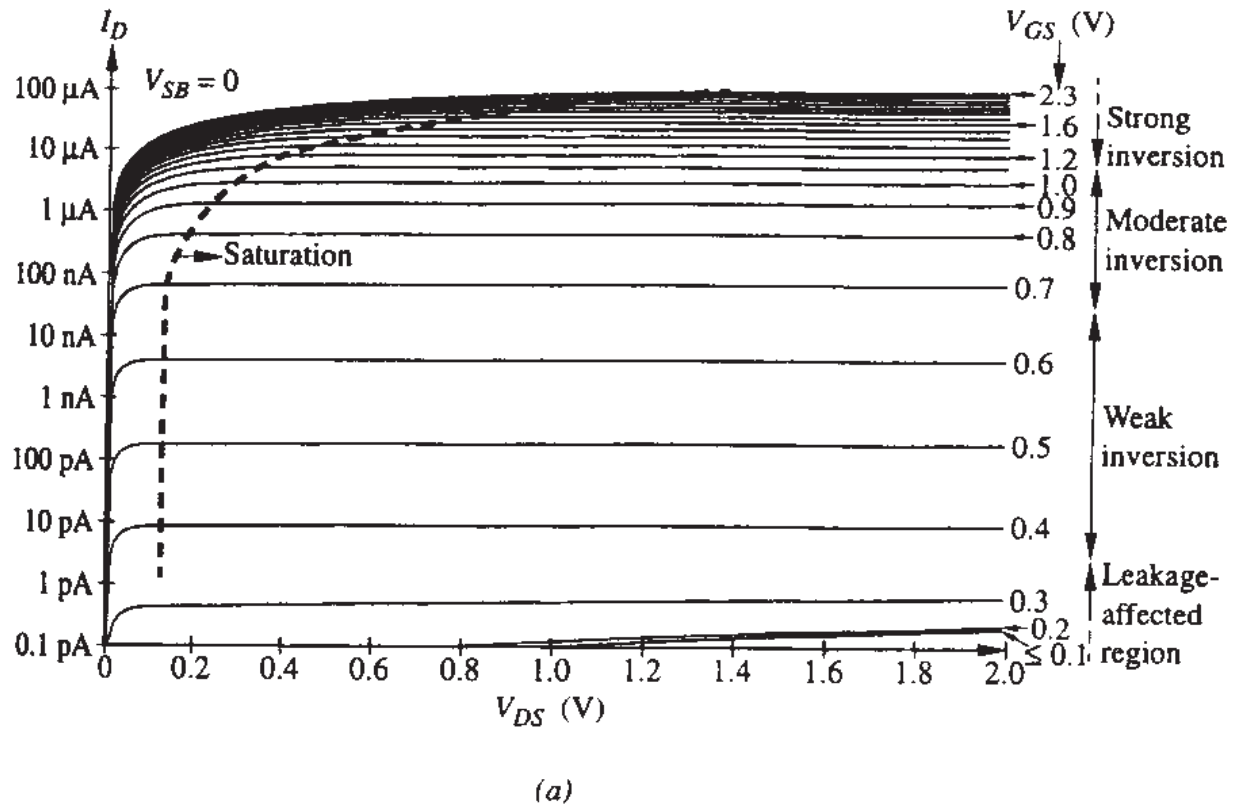


FIGURE 1.27

I_D - V_{DS} characteristics with V_{GS} as a parameter for a specific device in the connection of Fig. 1.22. (a) Logarithmic I_D axis; (b) linear I_D axis. (Figure reprinted with permission from Y. Tsividis, *Mixed Analog-Digital VLSI Devices and Technology*, McGraw-Hill, 1996.)

boundary between the regions of nonsaturation and saturation is indicated by a broken line; saturation is to the right of that line.

The same characteristics are shown in Fig. 1.27*b* using a linear I_D axis. Obviously this type of plot does not do justice to weak inversion. This region was in fact unknown for several years, until suspicious “leakage currents” in dynamic memories prompted researchers to take a more detailed look. The weak inversion region is now rather well understood and is an important region of operation for many applications, especially those for which very small power dissipation, or low voltage operation, is desired.

The above plots have been obtained for a long-channel device. For this device, the increase of I_D with V_{DS} in saturation is not easily seen on the scale used. For devices with short channels, the effect of V_{DS} on I_D is more evident. Devices with short channels will be considered in detail later in this book.

A careful look at transistor characteristic curves reveals a distinctly different behavior in each region of inversion. Let us begin with the weak inversion region, which is marked in Fig. 1.27*a*. In this region, the current turns out to be due to the diffusion of carriers (Sec. 1.3.3) and corresponds to the case of Fig. 1.26*a* and *b*. The spacing of consecutive curves in Fig. 1.27*a* can be seen to be almost equal, for equal V_{GS} increments. Given that the vertical axis is logarithmic, this means that I_D is, for all practical purposes, exponentially related to V_{GS} in weak inversion (Prob. 1.14). This behavior reminds one of the type of dependence of collector current on base-emitter voltage in a bipolar transistor. Typically, the range of V_{GS} values over which weak inversion behavior is observed can be a few tenths of a volt; nevertheless, because of the exponential behavior in this region, this is sufficient for a drain current variation of 3 or 4 orders of magnitude at room temperature.

In strong inversion, marked in Fig. 1.27*a* and *b*, the current turns out to be due to drift (Sec. 1.3.2). This mode of operation corresponds to the case in Fig. 1.25. In the saturation part of strong inversion, I_D is found to be approximately quadratic in V_{GS} for long-channel devices. This can be verified from the spacing of consecutive curves in Fig. 1.27*b* (Prob. 1.14).

In moderate inversion (marked in Fig. 1.27*a* and *b*), both drift and diffusion currents contribute significantly. This situation corresponds to Fig. 1.26*c*. In this region, I_D is neither exponential nor polynomial; rather its behavior changes gradually from one form of functional dependence to the other, as V_{GS} is raised by a few tenths of a volt. Within this region, the current can vary by a couple of orders of magnitude.

1.7 A BRIEF OVERVIEW OF THIS BOOK

In order to derive quantitative results for MOS transistor behavior, the study of many phenomena becomes necessary. In the following chapters, we will make clear which phenomena are responsible for which behavior, by studying them separately.

In Chap. 2, we will consider only the middle part of the transistor in Fig. 1.20, i.e., a gate-oxide-body structure with only gate and body terminals. Several

fundamental mechanisms and results will be studied for this structure. In Chap. 3, we will add one more terminal (corresponding to the source in Fig. 1.24). In the resulting three-terminal structure, the third terminal can be used to contact the inversion layer. We will consider the additional phenomena that occur because of the presence of this third terminal, notably the body effect. These two chapters cover phenomena whose manifestation does not rely on the presence of current flow; this approach makes the study of these phenomena easy. Nevertheless, the phenomena studied are also present when current is flowing, and thus the results derived in Chaps. 2 and 3 will be found to be directly applicable to the cases discussed in the rest of the book.

In Chap. 4, we will add the fourth terminal (the drain), which will make possible the flow of current in the complete transistor. The transistor will be assumed to be of large dimensions and on a uniform substrate. It will be seen that one-dimensional and pseudo-one-dimensional approaches are sufficient for this case. A fundamental approach to deriving expressions for the drain current will be discussed, one that results in a model valid in all regions of operation. Simpler, approximate models valid in particular regions will also be considered, as will several refinements.

In Chap. 5, we will allow for substrate nonuniformity, made possible through the process of ion implantation, and we will discuss how our models so far need to be modified in the presence of such nonuniformity.

In Chap. 6, we will consider what happens when the dimensions of the transistor are made small. We will see how our description of the various phenomena, and the resulting models, must be modified for small devices. Here we will see the limits of some analytical approaches, and we will explain the need for two-dimensional and pseudo-two-dimensional approaches.

Up to the end of Chap. 6, our interest will be to study the dc drain current as a function of the terminal voltages, the latter being assumed constant. In Chap. 7, we will allow those voltages to become functions of time, and we will consider the resulting varying charges and terminal currents. Such dynamic behavior cannot be described adequately by the results of the previous chapters. Depending on the speed of variation of the terminal voltages, we will see that models of differing complexity become possible. We remark that results in this and the previous chapters deal with *total* voltages and currents only.

The next two chapters deal with the relation of *small increments* of the terminal voltages to the resulting increments of the terminal currents. This “small-signal” behavior is of importance in the design not only of analog circuits, but also of digital circuits of high performance. Chapter 8 discusses approaches which are simple and satisfactory at low and medium frequencies, whereas Chap. 9 discusses additional considerations which become necessary at very high frequencies.

The last chapter of the book, Chap. 10, considers the process of “compact” modeling in which the several phenomena discussed separately must be put together in order to describe the transistor in a variety of situations. Such models are indispensable in computer-aided circuit design. Several considerations in the development of compact models, as well as ways to evaluate such models, will be discussed.

REFERENCES

1. R. B. Adler, A. C. Smith, and R. L. Longini, *Introduction to Semiconductor Physics*, SEEC Series, vol. 1, John Wiley, New York, 1964.
2. P. E. Gray, D. DeWitt, A. R. Boothroyd, and J. F. Gibbons, *Physical Electronics and Circuit Models of Transistors*, SEEC Series, vol. 2, John Wiley, New York, 1964.
3. P. E. Gray and C. L. Searle, *Electronic Principles: Physics, Models, and Circuits*, John Wiley, New York, 1969.
4. A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley, New York, 1967.
5. S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.
6. R. S. Muller and T. Kamins, *Device Electronics for Integrated Circuits*, John Wiley, New York, 1986.
7. M. Shur, *Introduction to Electronic Devices*, John Wiley, New York, 1996.
8. B. G. Streetman, *Solid State Electronic Devices*, Prentice-Hall, Upper Saddle River, N.J., 1995.
9. E. S. Yang, *Microelectronic Devices*, McGraw-Hill, New York, 1988.
10. J. L. Moll, *Physics of Semiconductors*, McGraw-Hill, New York, 1964.
11. R. F. Pierret, *Advanced Semiconductor Fundamentals*, Addison-Wesley, Reading, Mass., 1987.
12. G. W. Neudeck, *The PN Junction Diode*, Addison-Wesley, Reading, Mass., 1989.
13. W. E. Beadle, J. C. C. Tsai, and R. D. Plummer (editors), *Quick Reference Manual for Silicon Integrated Circuit Technology*, Wiley-Interscience, New York, 1985.
14. B. E. Deal, E. H. Snow, and C. A. Mead, "Barrier energies in metal-silicon dioxide-silicon structures," *Journal of Physics and Chemistry of Solids*, vol. 27, p. 1873, 1966.
15. S. Kar, "Determination of Si-metal work function differences by MOS capacitance technique," *Solid-State Electronics*, vol. 18, pp. 169–181, 1975.
16. E. H. Rhoderick, *Metal-Semiconductor Contacts*, Oxford University Press, London, 1978.
17. J. E. Lilienfeld, US Patents 1,745,175 (1930, filed October 26, 1926), 1,877,140 (1932, filed December 8, 1928) and 1,900,018 (1933, filed March 28, 1928).
18. O. Heil, British Patent 439,457 (filed and granted 1935).
19. W. Shockley and G. L. Pearson, "Modulation of conductance of thin films of semiconductors by surface charges," *Physical Review*, vol. 74, pp. 232–233, July 1948.
20. D. Kahng and M. M. Atalla, "Silicon-silicon dioxide field induced devices," *Solid-State Device Research Conference*, Pittsburgh, June 1960.
21. E. H. Snow, B. E. Deal, A. S. Grove, and C. T. Sah, "Ion transport phenomena in insulating films," *Journal of Applied Physics*, vol. 36, pp. 1665–1673, May 1965.
22. H. K. J. Ihantola, "Design theory of surface field-effect transistor," Stanford Electronics Laboratories, *Technical Report No. 1161-1*, Stanford University, California, September 1961.
23. S. R. Hofstein and F. P. Heinman, "The silicon insulated-gate field effect transistor," *Proceedings of the IEEE*, vol. 51, pp. 1190–1202, September 1963.
24. H. K. J. Ihantola and J. L. Moll, "Design theory of a surface field-effect transistor," *Solid-State Electronics*, vol. 7, pp. 423–430, June 1964.
25. C. T. Sah, "Characteristics of the metal-oxide-semiconductor transistors," *IEEE Transactions on Electron Devices*, vol. ED-11, pp. 324–345, July 1964.
26. C.-T. Shah, "Evolution of the MOS transistor—From conception to VLSI," *Proceedings IEEE*, vol. 76, no. 10, pp. 1280–1326, October 1988.
27. S. M. Sze, *VLSI Technology*, McGraw-Hill, New York, 1988.
28. W. S. Ruska, *Microelectronics Processing*, McGraw-Hill, New York, 1987.
29. S. K. Ghandi, *VLSI Fabrication Principles—Silicon and Gallium Arsenide*, 2d ed., Wiley, New York, 1994.
30. Y. P. Tsividis, *Mixed Analog-Digital Devices and Technology—An Introduction*, McGraw-Hill, New York, 1996.
31. J. Y. Chen, *CMOS Devices and Technology for VLSI*, Prentice-Hall, Englewood Cliffs, 1990.
32. J. P. Colinge, *Silicon-on-Insulator Technology: Materials to VLSI*, Kluwer Academic Publishers, Boston, 1991.

33. D. A. Grant and J. Gowar, *Power MOSFETS—Theory and Application*, John Wiley & Sons, New York, 1989.
34. C. H. Sequin, "A fluid model for visualizing MOS transistor behavior," Sec. 1.15 in C. Mead and L. Conway, *Introduction to VLSI Systems*, Addison-Wesley, Reading, Mass., 1980, pp. 29–33.
35. C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley, Reading, Mass., 1989.

PROBLEMS

- 1.1. Calculate the electron and hole concentrations in p -type Si with $N_A = 10^{17} \text{ cm}^{-3}$ at $T = 280 \text{ K}$, 300 K , and 330 K .
- 1.2. This problem is intended to give an idea of the order of magnitude involved for the various quantities discussed in Sec. 1.3. Consider the bar of Fig. 1.5 with $a = 100 \text{ } \mu\text{m}$, $b = 10 \text{ } \mu\text{m}$, $c = 2 \text{ } \mu\text{m}$, and n -type doping with concentration of 10^{15} cm^{-3} ; $V = 1 \text{ V}$. Find the value of the conductivity, the mobility, the conductance, the sheet resistance, the total mobile charge, the mobile charge per unit area, the field intensity, the drift velocity, the transit time, and the current.
- 1.3. Prove that (1.3.23) is valid for the case of Fig. 1.8b.
- 1.4. (a) Prove mathematically that (1.3.14) and (1.3.15) are valid even if the electron concentration varies with depth as long as it is uniform horizontally.
(b) Prove mathematically that (1.3.20) and (1.3.22) are valid for the conditions stated in the paragraph following (1.3.20).
- 1.5. Verify (1.4.1).
- 1.6. Calculate the contact potential of Au to n -type Si with $N_D = 10^{16} \text{ cm}^{-3}$ at 300 K .
- 1.7. A voltmeter with both of its leads made out of a metal X measures a voltage V across a battery. Show that if one of the voltmeter's leads is replaced by a different material Y , the reading will not be affected.
- 1.8. Assume that the two terminals of a voltage source are made out of different material. Show that (1.4.11) and (1.4.12) are valid if V_{source} is defined as the voltage measured by an ideal voltmeter when it is attached to the terminals of the source.
- 1.9. For a two-sided step pn junction with neither of its sides degenerate, prove that the built-in potential ϕ_{bi} is given by $\phi_i \ln(N_A N_D / n_i^2)$.
- 1.10. Plot the junction capacitance vs. reverse-bias voltage (from 0 to 5 V) for a silicon n^+p junction of area $200 \text{ } \mu\text{m}^2$ with $N_A = 5 \times 10^{16} \text{ cm}^{-3}$. Assume $\phi_{Fn} = -0.56 \text{ V}$.
- 1.11. Consider a semiconductor in equilibrium. Express the fact that the *total* current (drift plus diffusion components) must be zero, by using (1.3.11a) (in the footnote) and (1.3.17). Applying (1.2.7), show that the Einstein relation (1.3.18) results.
- 1.12. Study the material on basic laws of electrostatics in Appendix B and provide detailed derivations for results (B.4) to (B.6) given there.
- 1.13. Study the material on pn junctions in Appendix C, and provide detailed derivations for the results given there with the help of the basic electrostatics laws from Appendix B. From these results find the charge per unit area on each side of the junction. Prove (1.5.7) for the case $N_D \gg N_A$.
- 1.14. (a) Show that, within the weak inversion saturation region marked in Fig. 1.27a, I_D is of the form $I_1 \exp(V_{GS}/V_1)$, and deduce the values for I_1 and V_1 .
(b) Show that within the strong inversion saturation region marked in Fig. 1.27b, I_D is of the form $k_1(V_{GS} - V_T)^2$, and deduce the values for k_1 and V_{GS} .

CHAPTER 2

THE TWO- TERMINAL MOS STRUCTURE

2.1 INTRODUCTION

In our gradual development toward the MOS transistor, we consider in this chapter the two-terminal MOS structure. This structure is often referred to as *MOS capacitor* and is shown in Fig. 2.1. The acronym *MOS*, standing for metal-oxide-semiconductor, is used independently of whether the gate is actually made of metal or whether the insulator is silicon dioxide. In the days when gates were made of metal, it became common to use the term *semiconductor* to refer unambiguously to the substrate material. Such usage continues today, even though the gate is normally made of polysilicon.

The two-terminal MOS structure has been the subject of extensive studies over many years. Such studies have resulted in a detailed understanding of the structure and identified the sources of a number of undesirable effects that plagued early work. This led to the development of better fabrication methods which greatly reduced such effects and which made possible MOS transistors with high performance. A detailed study of the two-terminal MOS structure and a related history can be found elsewhere.¹ Here we will only discuss those aspects directly relevant to the objective of this book. We will consider the various potentials and charges developed in the two-terminal MOS structure (assumed made with a modern fabrication process) when a voltage is applied between gate and substrate. We will also consider the capacitance properties of the structure.

Most of the foundation for our later study of the complete MOS transistor is laid in this chapter. To lay this foundation adequately, a considerable amount of cross-

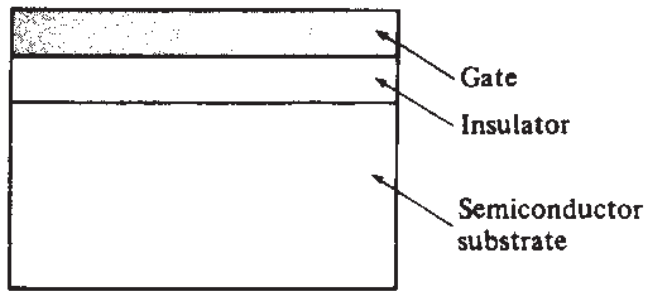


FIGURE 2.1
A two-terminal MOS structure.

referencing and page turning will be necessary, in order to properly derive several results and relate them to each other. We ask for the reader's patience in this respect.

2.2 THE FLAT-BAND VOLTAGE

We begin our discussion with an "academic" case, shown in Fig. 2.2a. Here we assume that the gate is made of the same crystalline material as the substrate (in this case *p*-type silicon), with the same doping concentration. We also assume that somehow the same material is used to connect the gate to the substrate, in a sense making one an extension of the other. No charges are shown in the silicon, since to each positively charged hole of the *p*-type material there corresponds a negatively charged acceptor atom, from which the hole has originated, and there is no reason for the holes to pile up in any particular region. The charges, therefore, cancel and the material is shown as electrically neutral everywhere. The structure of Fig. 2.2a is basically symmetric; assuming no parasitic charges have been introduced during fabrication, no field can exist in the insulator or the semiconductor, and there is no reason for carriers to be attracted toward the insulator-substrate interface.

Consider now a realistic case, as shown in Fig. 2.2b. The gate is made out of a certain material, not necessarily the same as the substrate material. A metal is used to contact the gate material and form the *gate terminal G* as shown. In the following, the subscript *G* will be used to indicate the *gate terminal* rather than the gate itself. The body (substrate) is contacted through a back metal plate as shown. This plate, in turn, is contacted through some metal, which thus forms the *body terminal B*. Again, the subscript *B* will be used to denote the *body terminal*, as shown, rather than the body itself.

Let us now short-circuit the gate terminal to the body terminal by using a wire as shown. Consider going from the gate material through the external connection to the bulk. In doing so we encounter several contact potentials. However, from our discussion in Sec. 1.4 we know that their sum will only depend on the first and the last material, being independent of any material in between. Applying (1.4.7) we have

$$\begin{array}{l} \text{Sum of all contact potentials} \\ \text{from gate, through external} \\ \text{connection, to bulk} \end{array} = \phi_{\text{gate material}} - \phi_{\text{bulk material}} \quad (2.2.1)$$

where each quantity in the right-hand side is characteristic of the corresponding material, as discussed in Sec. 1.4. The existence of a nonzero potential between the gate material and the bulk causes net charges to appear on both sides of the oxide. If, e.g., this potential is negative, the polarity of the charges will be as shown in Fig. 2.2b.

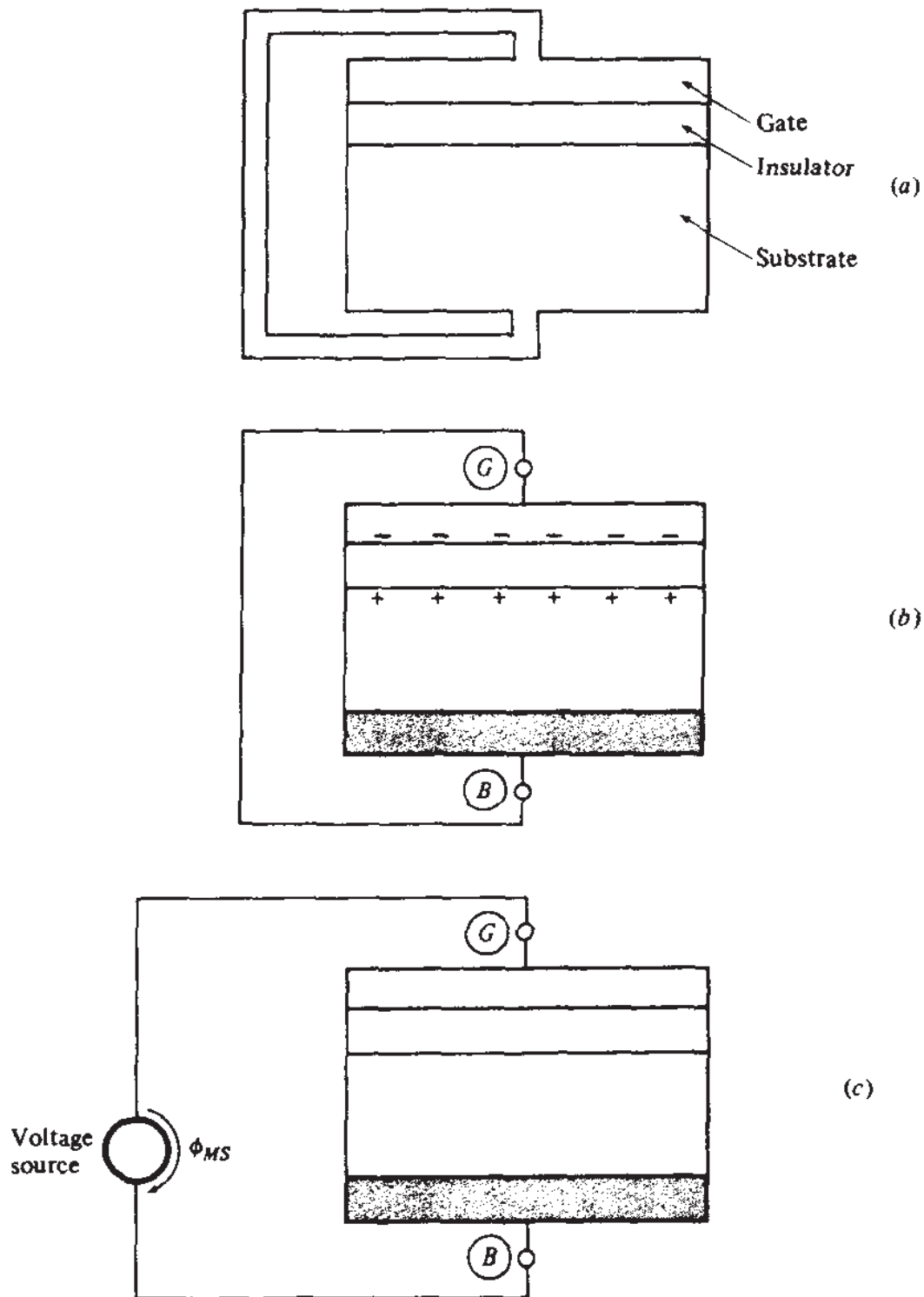


FIGURE 2.2

(a) A two-terminal MOS structure with gate, substrate, and short-circuiting external connection all made out of the same semiconductor material; (b) a MOS two-terminal structure (with gate and substrate made of different materials, and a bottom metal contact) with zero effective interface charge and with the gate terminal short-circuited to the substrate terminal; (c) the structure of (b) with a voltage source placed in the loop, so that the surface charge becomes zero; (d) effect of effective interface charge Q_o ; (e) the structure in (d) with additional external bias so that the surface charge becomes zero.

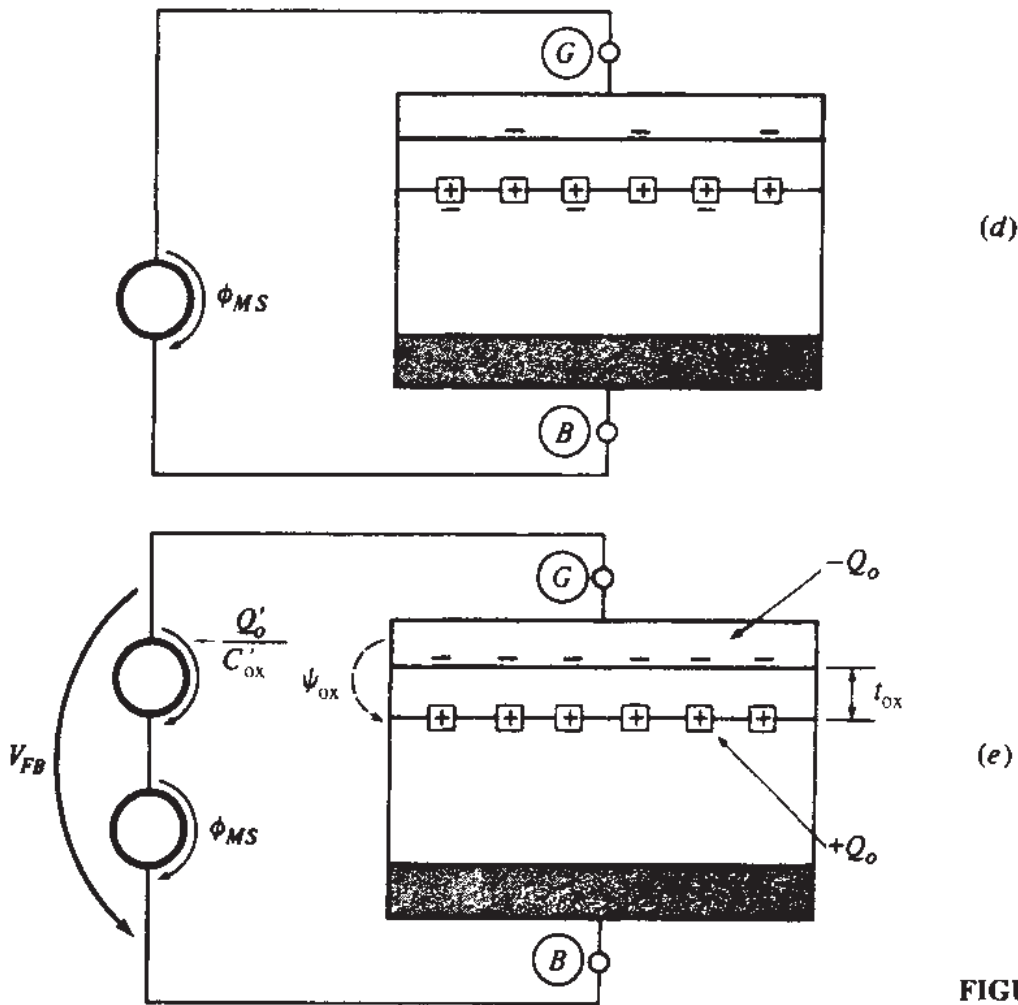


FIGURE 2.2 (continued).

Consider now the following question: Can an external voltage be applied in such a way that the net charges disappear? This case is shown in Fig. 2.2c. To make the total potential from the gate material through the external connection to the bulk equal to zero, the voltage of the external source must precisely cancel the sum of the contact potentials. From (2.2.1) we see that it must have a value given by

$$\phi_{MS} \equiv \phi_{\text{bulk material}} - \phi_{\text{gate material}} \quad (2.2.2)$$

where ϕ_{MS} is a widely used symbol; the value of ϕ_{MS} can be calculated from the above equation and Table 1.1.[†]

Example 2.1. For the common case of a degenerate polysilicon gate, we obtain, using (2.2.2) and Table 1.1:

For n^+ polysilicon gate: $\phi_{MS} = -\phi_F - 0.56\text{V}$

For p^+ polysilicon gate: $\phi_{MS} = -\phi_F + 0.56\text{V}$

[†]As expected from the footnote following (1.4.4), readers preferring to use "work functions" can write the right-hand side of (2.2.2) as $(1/q)(W_M - W_S)$, where W denotes work function and the subscripts M and S , which are commonly used, refer to the gate and semiconductor materials, respectively. Such usage is the origin of the subscripts MS in ϕ_{MS} .

The effect of contact potentials is not the only one that can cause a net concentration of charges in the substrate in the absence of external bias. Another cause is a "parasitic" charge that exists within the oxide as well as at the oxide-semiconductor interface. This charge consists of four parts¹:

1. An *oxide fixed charge* exists very close to the oxide-semiconductor interface due to the mechanisms of oxide formation at the time such formation is completed. This charge is found to be rather independent of oxide thickness, substrate doping type (n or p), and doping concentration.
2. A so-called *oxide trapped charge* can exist throughout the oxide, but usually close to either of its interfaces to the substrate or the gate. This charge can be acquired through radiation, photoemission, or the injection of high-energy carriers from the substrate.
3. A *mobile ionic charge* can exist within the oxide due to contamination by alkali ions (often sodium) introduced by the environment during fabrication. This charge can move within the oxide under the presence of an electric field.
4. An *interface trap charge* (also called *fast surface-state charge*) exists at the oxide-semiconductor interface. It is caused by defects at that interface, which give rise to charge "traps"; these can exchange mobile carriers with the semiconductor, acting as donors or acceptors.

The above charges initially inhibited the development of high-performance MOS devices because their nature was not well understood and they were difficult to control. In the early days, for example, the mobile ionic charge mentioned above was large and drifted about the oxide, depending on the externally applied voltage and the temperature. This charge, often introduced into the oxide during fabrication through the hands of the people handling the device, resulted in very uncontrollable characteristics. Today, after many years of work, the above four types of charge have been greatly reduced through appropriate fabrication techniques.¹ Throughout this book we assume that the devices we are dealing with have been fabricated by using such techniques. Until further notice, we will assume that all parasitic charge is located *at* the oxide-semiconductor interface and that its value, denoted by Q_o , is fixed.[†] Devices in which immobile charge exists *within* the oxide (uniformly distributed along the horizontal dimension in Fig. 2.2) can be modeled by assuming no such charge exists and by adjusting Q_o accordingly (Prob. 2.16); thus Q_o can be taken to be the *effective interface charge*. This charge is almost always positive for both p - and n -type substrates. In modern devices, the effective interface ion density is about 10^{10} to 10^{11} ions/cm², corresponding to an effective interface charge density of 1.6×10^{-9} to 1.6×10^{-8} C/cm².

[†]The case of devices in which a significant fraction of Q_o is not fixed but depends instead on the externally applied potential (due to the interface traps mentioned above) will be considered later.

The effective interface charge Q_o is shown inside the little squares in Fig. 2.2d. A battery of value ϕ_{MS} is used in this figure to cancel the effect of the contact potentials discussed previously, so that we can study the effect of Q_o by itself. The charge Q_o will cause a total charge $-Q_o$ to appear in the system as demanded by charge neutrality. As shown in Fig. 2.2d, part of that charge will appear at the gate and part in the semiconductor. If it is desired to eliminate the latter, one can note that, if all the required balancing charge $-Q_o$ were provided on the gate, no charge would be induced in the semiconductor. To provide a charge $-Q_o$ on the gate, a battery can be connected in series with the external circuit, with the (-) terminal toward the gate. The following is now clear from basic electrostatics (Appendix B): Since at the gate and substrate ends of the oxide we must have charge $-Q_o$ and Q_o , respectively, the potential drop across the oxide, ψ_{ox} , defined from the gate through the oxide to the substrate, must be equal to $-Q_o/C_{ox}$, where C_{ox} is the total capacitance between the two ends of the oxide. This is exactly the voltage that must be provided by the battery, as shown in Fig. 2.2e.

In terms of quantities per unit area Q'_o and C'_{ox} , we have, for ψ_{ox} in Fig. 2.2e.

$$\psi_{ox} = -\frac{Q'_o}{C'_{ox}} \quad (2.2.3)$$

The oxide capacitance per unit area is given by (Appendix B)

$$C'_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \quad (2.2.4)$$

where t_{ox} is the thickness of the insulator and ϵ_{ox} is its permittivity, given by

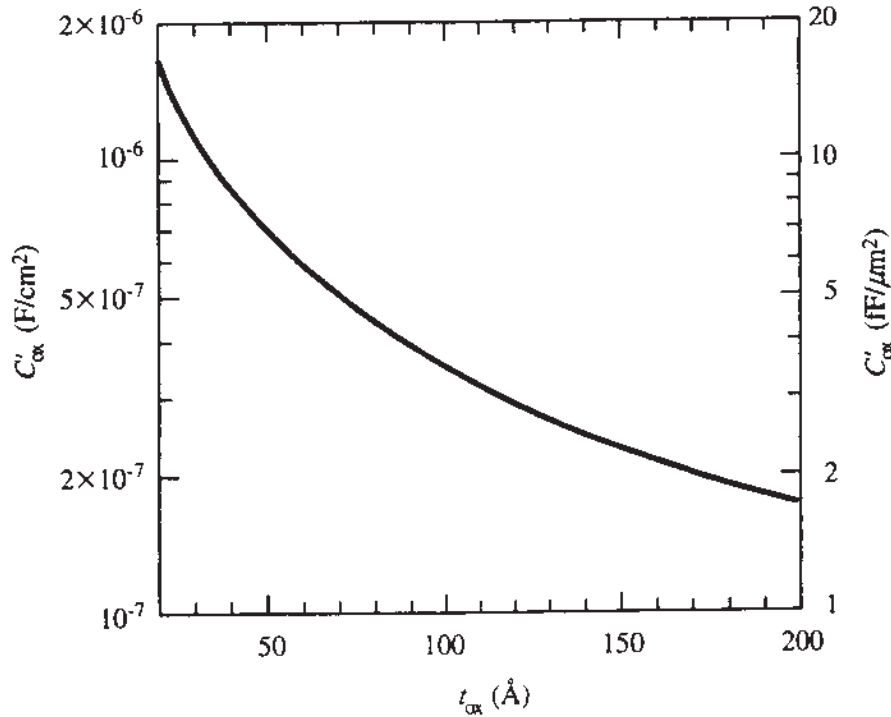
$$\epsilon_{ox} = k_{ox}\epsilon_0 \quad (2.2.5)$$

with ϵ_0 the permittivity of free space (8.854×10^{-14} F/cm) and k_{ox} the dielectric constant of the insulator; for SiO_2 , $k_{ox} = 3.9$. Equation (2.2.4) is plotted in Fig. 2.3.

We have therefore seen that an external voltage can be used between the gate and substrate terminals to keep the semiconductor everywhere neutral by canceling the effects of the contact potentials and Q'_o . This voltage is called the *flat-band voltage*[†] and is denoted by V_{FB} . From Fig. 2.2e we have the expression for the flat-band voltage:

$$V_{FB} = \phi_{MS} - \frac{Q'_o}{C'_{ox}} \quad (2.2.6)$$

[†]The name *flat-band* originates from an alternative description of semiconductor phenomena, which uses the concept of energy bands. It turns out that, for the uniform substrate assumed, those bands are "flat" in the semiconductor for the situation illustrated in Fig. 2.2e, and that the electric field throughout the body material is zero (Appendix D).

**FIGURE 2.3**

Capacitance per unit area vs. thickness for a silicon dioxide insulator.

Example 2.2. Calculate the flat-band voltage for a p -type substrate with $N_A = 9 \times 10^{16} \text{ cm}^{-3}$, an SiO_2 insulator with a thickness $t_{ox} = 100 \text{ Å}$, and an n -type polysilicon gate with $N_D = 10^{20} \text{ cm}^{-3}$. The interface charge Q'_o is 10^{-8} C/cm^2 .

The Fermi potential of the substrate is, from (1.4.2), equal to 0.41 V . From (2.2.2) and Table 1.1 or directly from Example 2.1, we have

$$\phi_{MS} = -0.41 \text{ V} - 0.56 \text{ V} = -0.97 \text{ V}$$

From (2.2.4), $C'_{ox} = 3.453 \times 10^{-7} \text{ F/cm}^2$, and therefore $-Q'_o/C'_{ox} = -0.03 \text{ V}$. Therefore, from (2.2.6),

$$V_{FB} = -0.97 \text{ V} - 0.03 \text{ V} = -1 \text{ V}$$

2.3 POTENTIAL BALANCE AND CHARGE BALANCE

We will now discuss how the substrate is affected when the externally applied voltage V_{GB} assumes values different from the flat-band voltage V_{FB} . Consider as an example a MOS structure with a p -type substrate, shown in Fig. 2.4a. An arbitrary value of V_{GB} will in general cause charges to appear in the semiconductor. Practically all of these charges will be contained within a region adjacent to the top surface of the semiconductor, which is shown shaded in Fig. 2.4a. Outside this region the substrate is practically neutral. We define the *surface potential* ψ_s as the total potential drop

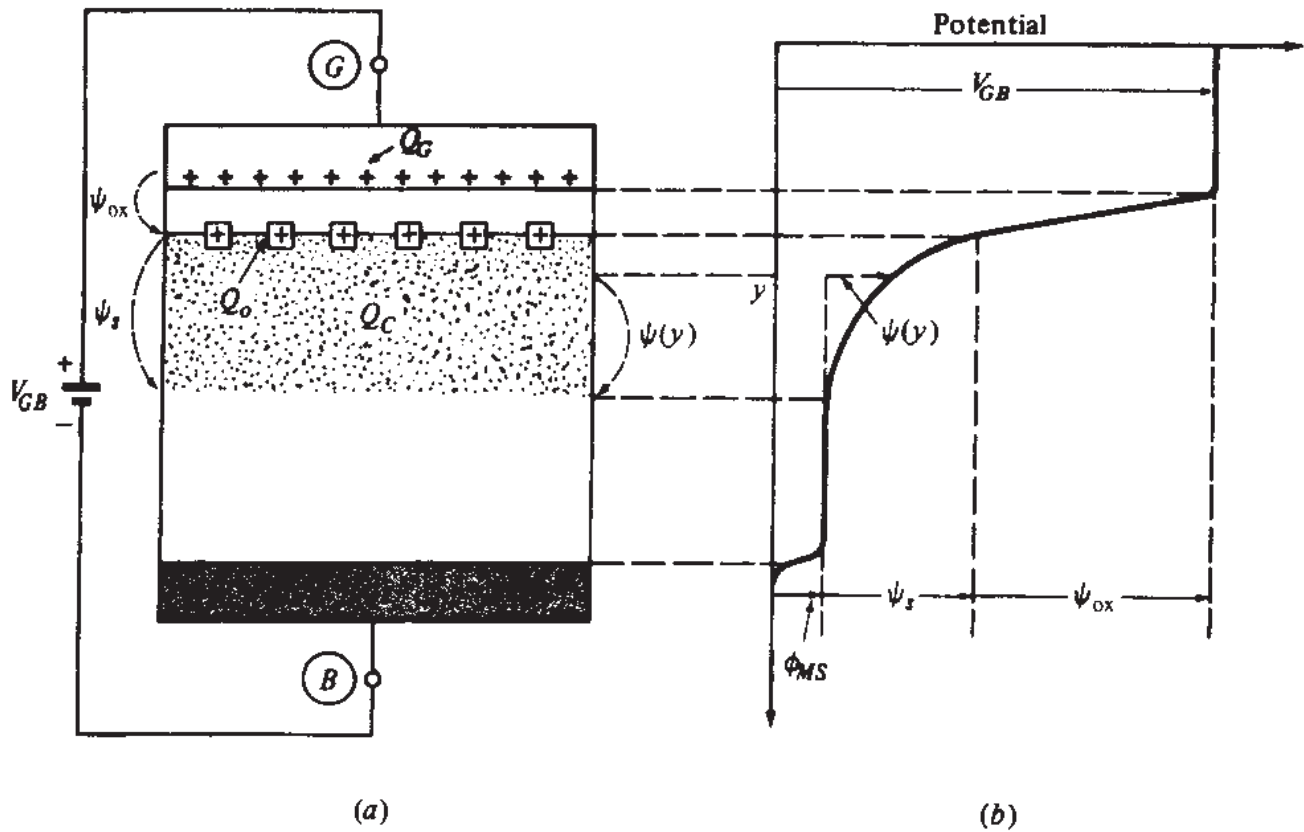


FIGURE 2.4

(a) A p -substrate two-terminal MOS structure under general gate-substrate bias; (b) potential distribution assuming the gate, the substrate cap, and the external wires are all made of the same material (the special case of $\psi_s > 0$ has been assumed in drawing this plot).

across the region, defined from the *surface* to a point in the *bulk* outside that region.[†] We will be more specific about the width of the region later on.

Four kinds of potential drops are encountered in the loop, as seen in Fig. 2.4:

1. The voltage of the external source V_{GB} .
2. The potential drop across the oxide ψ_{ox} .
3. The surface potential ψ_s .
4. Several contact potentials. Their sum, when going *clockwise*, is ϕ_{MS} , as seen from (2.2.1) and (2.2.2).

Going around the loop, we can write:

$$\boxed{V_{GB} = \psi_{ox} + \psi_s + \phi_{MS}} \quad (2.3.1)$$

[†]The top surface of the semiconductor is commonly referred to simply as “the surface” in MOS literature. The term *surface potential* is widely used as above.²⁻⁶ However, in some treatments,⁷ this term is used for a quantity which differs from our ψ_s by the Fermi potential ϕ_F .

The “potential balance” expressed by the above equation is illustrated in Fig. 2.4b for the simple case in which the gate, the substrate contact, and the wires are all assumed made of the same material (e.g., Al); then, the only contact potential involved is between the substrate and its metal cap. If more than one contact is involved, (2.3.1) still holds, but a potential plot should then include all contact potentials, adding up to the value given by (2.2.2).

Note that in (2.3.1) ϕ_{MS} is a known constant; therefore, any changes in V_{GB} must be balanced by changes in ψ_{ox} and ψ_s :

$$\Delta V_{GB} = \Delta\psi_{ox} + \Delta\psi_s \quad (2.3.2)$$

Consider now the charges in the system. In Fig. 2.4 we encounter three kinds of charges:

1. The charge on the gate Q_G .
2. The effective interface charge Q_o .
3. The charge in the semiconductor under the oxide Q_C .

These charges must balance one another for overall charge neutrality in the system:

$$Q_G + Q_o + Q_C = 0 \quad (2.3.3)$$

or, in terms of charges per unit area,

$$Q'_G + Q'_o + Q'_C = 0 \quad (2.3.4)$$

From now on, we will be using *charges per unit area* rather than charges most of the time. For brevity, though, we will often omit the words “per unit area”; these will be implied in the context.

Notice that if Q'_G is changed, the balance required by the foregoing equation will be achieved through a change in Q'_C , since the equivalent interface charge Q'_o is for the present assumed fixed:

$$\Delta Q'_G + \Delta Q'_C = 0 \quad (2.3.5)$$

The potential balance equation (2.3.1) and the charge balance equation (2.3.4) have been deduced from general fundamental physical laws. As we will see, particular properties of the MOS system will impose additional relations between the quantities appearing in the above equations.

2.4 EFFECT OF GATE-SUBSTRATE VOLTAGE ON SURFACE CONDITION

Let us now consider the effect of V_{GB} on the condition of the region containing Q_C in Fig. 2.4. A *p*-type substrate is assumed. Depending on whether V_{GB} is equal to, less than, or greater than the flat-band voltage V_{FB} , three cases are distinguished below.

2.4.1 Flat-Band Condition

This case has already been discussed in detail in Sec. 2.2 and is illustrated in Fig. 2.2e. From the associated discussion we have

$$V_{GB} = V_{FB} \quad (2.4.1)$$

$$Q'_C = 0 \quad (2.4.2)$$

$$\psi_s = 0 \quad (2.4.3)$$

2.4.2 Accumulation

Consider the case in which V_{GB} decreases below V_{FB} (e.g., for the device of Example 2.2, this means that V_{GB} is *more negative* than -1 V). The negative change of V_{GB} (relative to flat band) will cause a negative change in Q'_G which, according to (2.3.5), must be balanced by a positive change in Q'_C above the value given by (2.4.2). Thus, holes will accumulate at the surface to provide a net positive charge. This condition is called *accumulation* and is illustrated in Fig. 2.5a. The negative change in V_{GB} will be shared by negative changes in ψ_{ox} and ψ_s , and (2.3.1) will remain valid. In accumulation, therefore, we have

$$V_{GB} < V_{FB} \quad (2.4.4)$$

$$Q'_C > 0 \quad (2.4.5)$$

$$\psi_s < 0 \quad (2.4.6)$$

2.4.3 Depletion and Inversion

Assume now the case in which V_{GB} increases above V_{FB} . The total charge on the gate Q_G will become more positive than the value in flat band (that value is $-Q_o$ in Fig. 2.2e). An example is shown in Fig. 2.5b, where it is assumed that the resulting Q'_G is positive. The positive change in Q'_G (relative to flat band) must be balanced by a negative change in Q'_C so that (2.3.3) remains valid. Also, the positive change in V_{GB} will be shared among ψ_{ox} and ψ_s , with (2.3.1) remaining valid. Accordingly, we have:

$$V_{GB} > V_{FB} \quad (2.4.7)$$

$$Q'_C < 0 \quad (2.4.8)$$

$$\psi_s > 0 \quad (2.4.9)$$

Let us now consider the nature of the negative charge Q'_C . If V_{GB} is not much higher than V_{FB} , the positive potential at the surface with respect to the bulk will simply drive holes away from the surface, leaving it depleted. This condition is called *depletion* and it is illustrated in Fig. 2.5b. More precisely, as V_{GB} is raised above V_{FB} ,

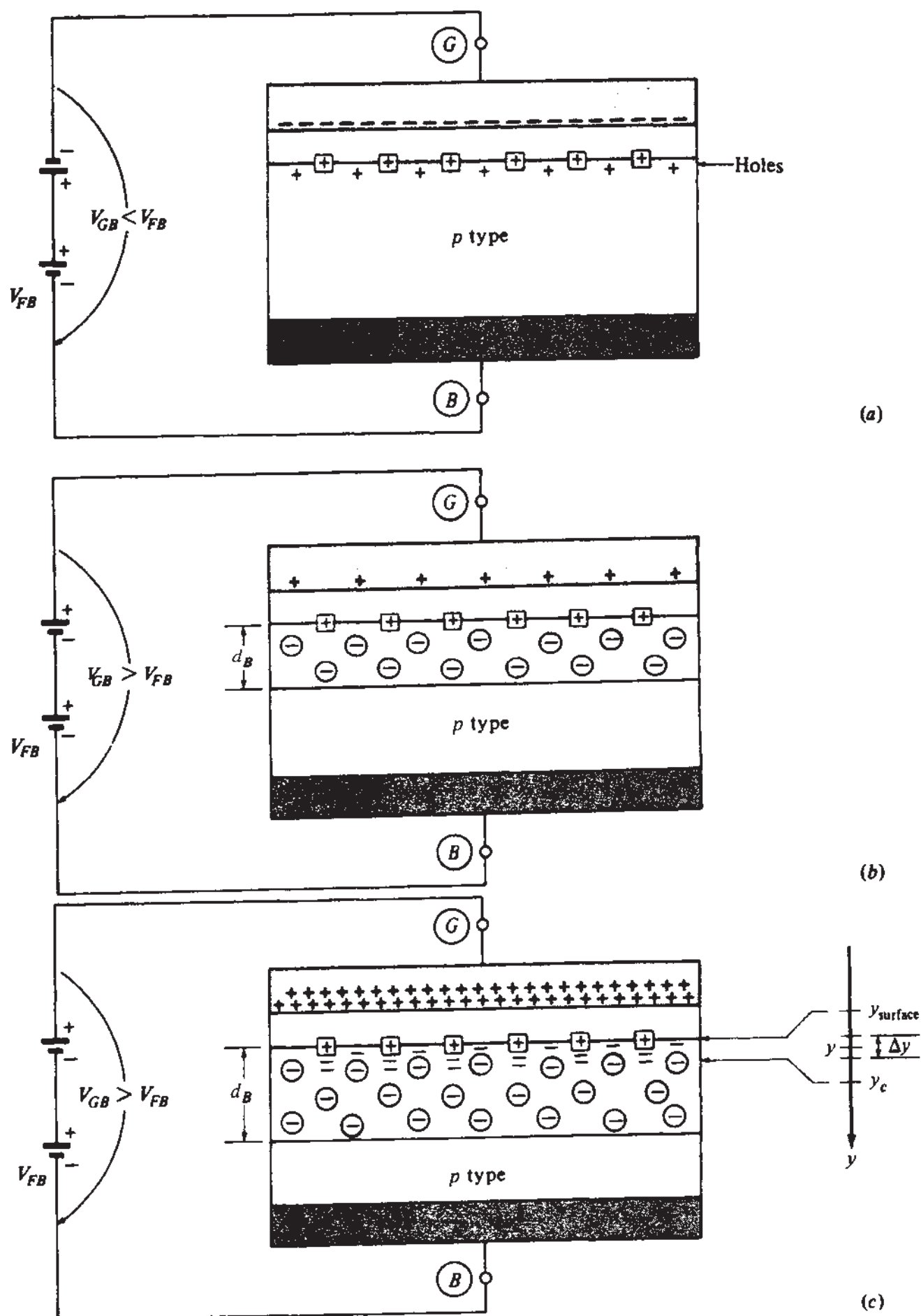


FIGURE 2.5

A MOS two-terminal structure in (a) accumulation; (b) depletion; and (c) inversion.

the hole density will keep decreasing well below the doping concentration value N_A . For practical purposes, then, the charge Q_C is due to the uncovered acceptor atoms, each of which contributes a charge $-q$, and we can assume the presence of a "depletion region" as in the p side of the pn junction in Sec. 1.5.

As V_{GB} is increased further, more acceptor atoms are uncovered, and ψ_s becomes sufficiently positive to attract a significant number of free electrons to the surface; each of these electrons will also contribute a charge $-q$ to Q_C . Note that in the two-terminal MOS structure we are considering, these electrons come from the relatively slow process of electron-hole generation in the depletion region, caused by the thermal vibration of the lattice (assuming no radiation is present). Eventually, with a sufficiently high V_{GB} the density of electrons will exceed that of holes at the surface. This is a situation opposite from that normally expected in a p -type material; we now have surface *inversion*. The situation is illustrated in Fig. 2.5c.†

At this point we should warn the reader that the term *region* is often used with two different meanings in MOS work. Thus, for example, "depletion region" could refer to the physical region containing the ionized acceptor atoms in Fig. 2.5b; or it could refer to the region of V_{GB} (or ψ_s) values in which we have depletion as defined above. Which of the two meanings is employed will usually be clear from the context.

The oxide blocks the flow of current and equilibrium is maintained in the substrate. Deep in the neutral bulk outside the depletion region, the values of the hole and electron concentrations will be p_o and n_o , as given by (1.2.4) and (1.2.5). We can relate the electron concentration at the surface to that in the bulk by using (1.2.7):

$$n_{\text{surface}} = n_o e^{\psi_s/\phi_t} \quad (2.4.10)$$

Using (1.4.1a) and (1.4.1b), the above can be written as follows:

$$n_{\text{surface}} = n_i e^{(\psi_s - \phi_F)/\phi_t} \quad (2.4.11a)$$

$$= p_o e^{(\psi_s - 2\phi_F)/\phi_t} \quad (2.4.11b)$$

and, since $p_o \approx N_A$ from (1.2.4), we have

$$n_{\text{surface}} \approx N_A e^{(\psi_s - 2\phi_F)/\phi_t} \quad (2.4.12)$$

The surface electron concentration is plotted vs. the surface potential in Fig. 2.6. Some interesting points are shown in the figure. At $\psi_s = \phi_F$, n_{surface} becomes equal to the intrinsic concentration as seen from (2.4.11a); from (1.2.6) then, $n_{\text{surface}} = p_{\text{surface}}$. This is defined as the limit point between the depletion and inversion regions, as indicated in the figure. Of course, as seen from (2.4.11a), n_{surface} will be nonzero even in depletion but will be much smaller than n_i , even for ψ_s smaller than ϕ_F by only a few

†The application of the basic laws of electrostatics to produce plots of charge density, electric field, and potential for this case is illustrated in Appendix E.

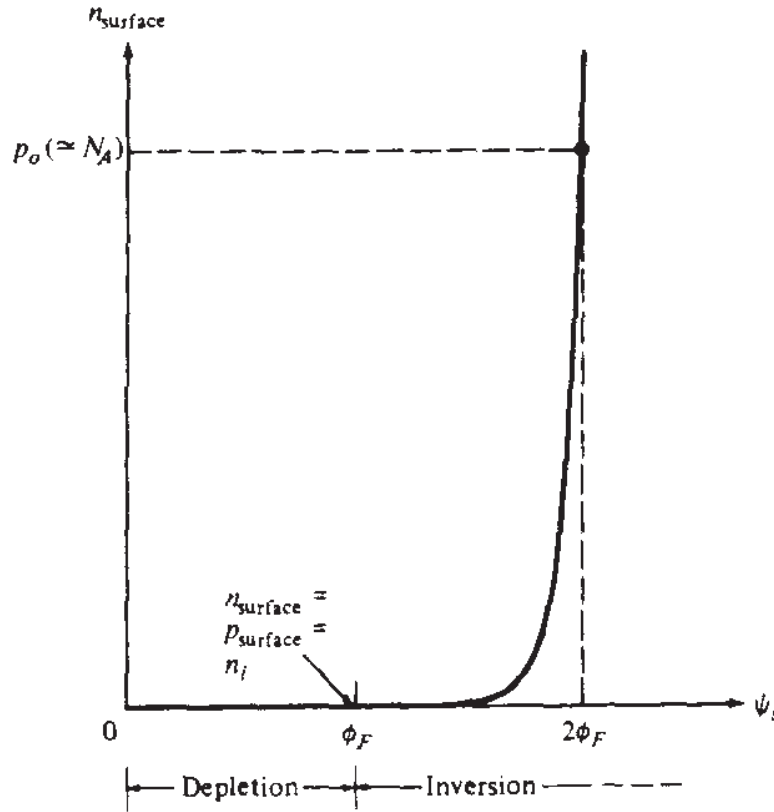


FIGURE 2.6

Electron concentration at the surface vs. surface potential. Linear axes are used.

ϕ_F . With increasing ψ_s above ϕ_F , n_{surface} increases drastically, and at $\psi_s = 2\phi_F$ we have $n_{\text{surface}} = p_o \approx N_A$, as seen from (2.4.11b) and (2.4.12).

If the substrate is made of n -type material, the above picture should be modified in a rather obvious manner. The inversion layer in this case will consist of holes, which will be attracted to the surface if V_{GB} is sufficiently negative. The immobile charge in the depletion region will consist of positively charged ionized donor atoms. With V_{GB} sufficiently positive, electrons will pile up at the surface and we will have accumulation. Since the case of n -type substrates is complementary to that of p -type substrates, it will not be discussed separately.

2.4.4 General Analysis

Of the three operation regions defined above, inversion is by far the most important for the purposes of this book. This region is treated in detail in the following section. It should be noted at this point, though, that a general analysis is possible through which one can determine ψ_s and Q'_C for any value of V_{GB} , be it in accumulation, depletion, or inversion. Although such general analysis will not be of much use in this book, we will summarize the principles behind it for completeness. The mathematical details will be left for Appendix F. Consider a point or ordinate y in the substrate (Fig. 2.4), and let $\psi(y)$ be the potential there with respect to the bulk. From (1.2.7) we will have, for the electron concentration at y ,

$$n(y) = n_o e^{\psi(y)/\phi_T} \quad (2.4.13)$$

which reduces to (2.4.10) at the surface. Similarly, for the hole concentration at y we will have, according to (1.2.11),

$$p(y) = p_o e^{-\psi(y)/\phi_t} \quad (2.4.14)$$

In the general case, one must allow for the presence of both electrons and holes below the oxide. Thus the charge density from (1.2.12) is

$$\rho(y) = q[p(y) - n(y) - N_A] \quad (2.4.15)$$

Substituting in this (2.4.13) and (2.4.14), we can write Poisson's equation (1.2.13) as follows:

$$\frac{d^2\psi}{dy^2} = \frac{-q}{\epsilon_s} \left(p_o e^{-\psi(y)/\phi_t} - n_o e^{\psi(y)/\phi_t} - N_A \right) \quad (2.4.16)$$

With the help of this equation and the three above it, and for a given ψ_s , one can determine $\psi(y)$, $n(y)$, $p(y)$, and $\rho(y)$, and the corresponding total charge per unit area in the substrate, Q'_C (Appendix F). Unfortunately, only Q'_C can be derived in closed form, the other quantities requiring a numerical evaluation. As shown in Appendix F, we obtain

$$Q'_C = \mp \sqrt{2q\epsilon_s N_A} \sqrt{\phi_t e^{-\psi_s/\phi_t} + \psi_s - \phi_t + e^{-2\phi_F/\phi_t} (\phi_t e^{\psi_s/\phi_t} - \psi_s - \phi_t)} \quad (2.4.17)$$

where the $-$ sign in front of the right side of the equation must be used with $\psi_s > 0$ (depletion or inversion), and the $+$ sign with $\psi_s < 0$ (accumulation).

The quantity $\sqrt{2q\epsilon_s}$ is the same quantity encountered in our discussion of the pn junction in Sec. 1.5 and has a value of $5.8 \times 10^{-16} \text{ C} \cdot \text{cm}^{-1/2} \cdot \text{V}^{-1/2}$ for silicon.

To complete our set of basic equations, note that the charge per unit area above the oxide, Q'_G , can be simply related to the potential across the oxide, ψ_{ox} , and the oxide capacitance per unit area, C'_{ox} , by (Prob. 2.11):

$$Q'_G = C'_{ox} \psi_{ox} \quad (2.4.18)$$

We have now derived four equations that completely characterize the MOS structure under our assumptions. These equations are (2.3.1), (2.3.4), (2.4.17), and (2.4.18), and contain four unknowns: ψ_{ox} , ψ_s , Q'_C , and Q'_G . This system of equations can be solved numerically, to provide the values of the above quantities for a given MOS structure (i.e., given ϕ_{MS} , Q'_o , and N_A) and given the externally applied voltage V_{GB} . With ψ_s known from this solution, one can determine $\psi(y)$, $n(y)$, $p(y)$, and $\rho(y)$, as already discussed.

The above analysis is complicated. The complexity is, in part, due to the fact that complete generality was sought, i.e., validity in accumulation, depletion, and inversion. In this book, properties in accumulation and depletion are needed only

for evaluating some parasitic effects associated with turned-off MOS transistors (Chap. 7). In contrast, inversion is responsible for current conduction in MOS transistors and is much more important to us. By focusing on this region, certain approximations become possible which simplify the analysis of the MOS structure, as will be seen in the following section.

2.5 INVERSION

2.5.1 General Relations and Regions of Inversion

For substrate doping concentrations usually encountered in MOS transistor work (10^{16} to 10^{18} cm^{-3}), the Fermi potential ϕ_F can have values approximately between $13\phi_t$ and $18\phi_t$; thus, the quantity $2\phi_F$ in (2.4.17) lies between $26\phi_t$ and $36\phi_t$. It is easy to see then that in inversion (Fig. 2.5c), where $\psi_s \geq \phi_F$, (2.4.17) can be approximated by the following†:

$$Q'_C = -\sqrt{2q\epsilon_s N_A} \sqrt{\psi_s + \phi_t} e^{(\psi_s - 2\phi_F)/\phi_t} \quad (2.5.1)$$

The total charge (per unit area) below the oxide is the sum of the charge due to the electrons in the inversion layer Q'_I and the charge due to the ionized acceptor atoms in the depletion region Q'_B :

$$Q'_C = Q'_I + Q'_B \quad (2.5.2)$$

We now undertake the evaluation of Q'_I and Q'_B . Consider first the electrons in the inversion layer. At any point of ordinate y in Fig. 2.5c, the electron concentration $n(y)$ will be given by (2.4.13). As one goes away from the surface, $\psi(y)$ decreases from ψ_s toward zero, and $n(y)$ decreases rapidly owing to its exponential dependence on $\psi(y)$. Hence, one can choose a point $y = y_c$ below which the electron concentration will be negligible. Practically all of the free electrons are then contained in a layer between $y = y_{\text{surface}}$ and $y = y_c$ (Fig. 2.5c). The number of electrons contained in a thin layer of thickness Δy , parallel to the surface and centered around y , will be $n(y)(A \Delta y)$, where A is the cross-sectional area as seen from the top. The total charge due to these electrons will be $(-q)n(A \Delta y)$. We can then express the charge due to all electrons in the inversion layer, denoted by Q_I , as follows:

$$Q_I = \int_{y_{\text{surface}}}^{y_c} (-q)n(y)(A \, dy) \quad (2.5.3)$$

Therefore, the inversion layer charge per unit area Q'_I will be given by

$$Q'_I = -q \int_{y_{\text{surface}}}^{y_c} n(y) \, dy \quad (2.5.4)$$

†Equation (2.5.1) can also be derived directly for the situation in Fig. 2.5c, starting from the assumption that no holes are present in the depletion region (Prob. 2.12). This equation is actually valid even in the upper part of depletion (i.e., for ψ_s larger than a few ϕ_t).

Evaluating Q'_I in this manner is a lengthy process and can be done through numerical integration^{1,3} (Appendix F). We will instead follow a widely used simplifying approach: We will determine a sufficiently accurate expression for Q'_B and then return to evaluate Q'_I from (2.5.2) and (2.5.1). Let us then concentrate on the depletion region. As in the case of the pn junction, we will consider this region as being defined by a sharp boundary at a depth d_B below the surface (Fig. 2.5c). The inversion layer is at the top of this region. Numerical calculations (Appendix F) show that most of the charge in this layer is concentrated very close to the surface (within a few hundred angstroms). Since the depth d_B of the depletion region is normally much larger, we will assume that the inversion layer is a sheet of negligible thickness.⁸⁻¹³ This has been called the *charge sheet approximation*¹³ and implies that practically all of the depletion region is free of electrons. For a negligible thickness, the potential drop across the inversion layer will also be negligible (Prob. 2.13), and we can assume that all of the surface potential ψ_s is dropped across the depletion region in the p -type substrate. We can thus relate d_B to ψ_s , just as we did for the n^+p junction in Sec. 1.5. We will assume that the mobile carrier concentrations are negligible in comparison to the acceptor concentration inside the depletion region; this has been called the *depletion approximation* in Sec. 1.5. Solving Poisson's equation (1.2.13) under this assumption results in an equation analogous to (1.5.12) (Appendix E and Prob. 2.14):

$$d_B = \sqrt{\frac{2\epsilon_s}{qN_A}} \sqrt{\psi_s} \quad (2.5.5)$$

Let Q'_B be the charge per unit area due to the uncovered acceptor atoms in the depletion region. Corresponding to (1.5.13) we have:

$$Q'_B = -\sqrt{2q\epsilon_s N_A} \sqrt{\psi_s} \quad (2.5.6)$$

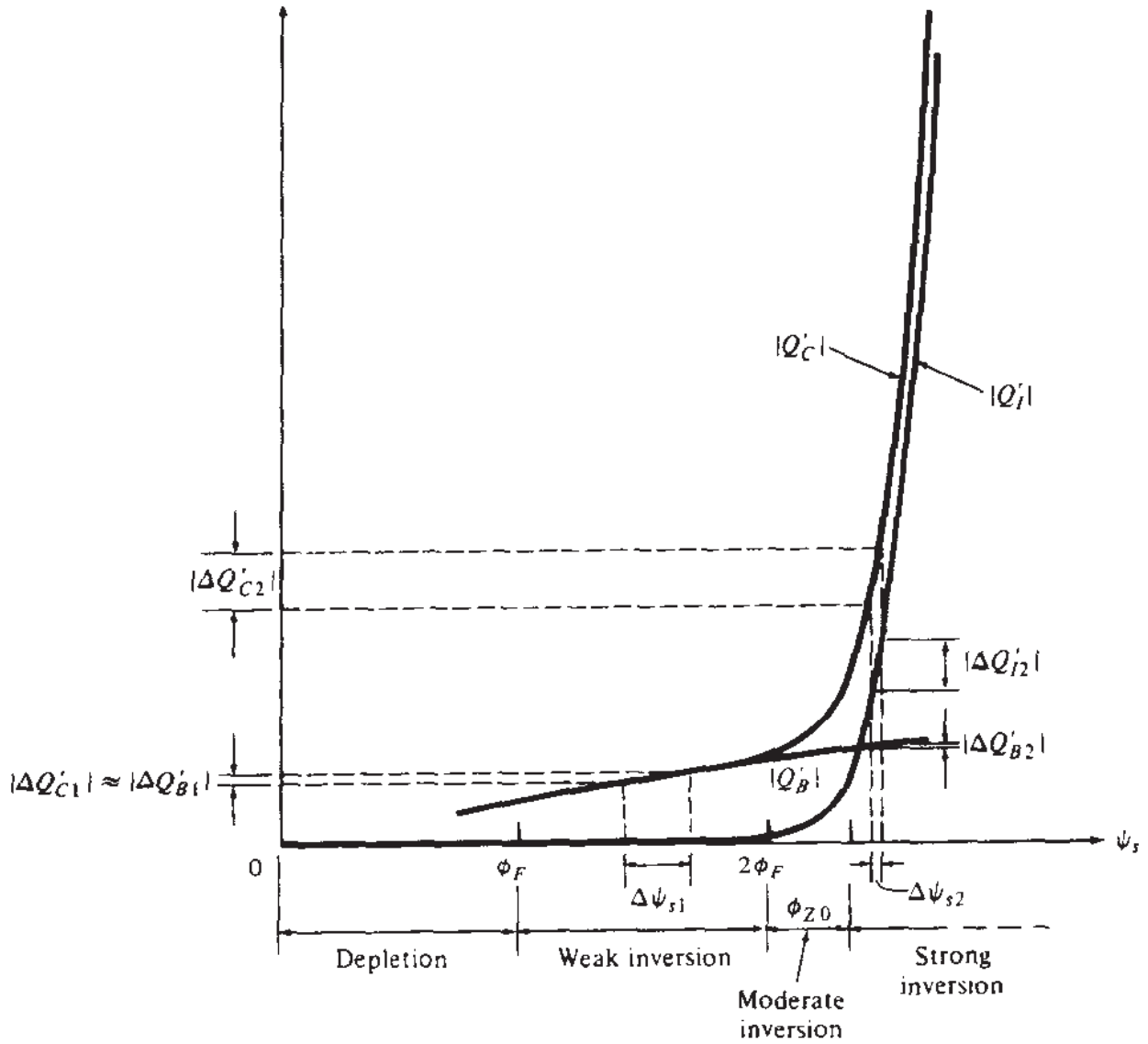
Using now the above equation and (2.5.1) in (2.5.2), we obtain the inversion layer charge per unit area†:

$$Q'_I = -\sqrt{2q\epsilon_s N_A} \left(\sqrt{\psi_s + \phi_i} e^{(\psi_s - 2\phi_F)/\phi_i} - \sqrt{\psi_s} \right) \quad (2.5.7)$$

We have plotted $|Q'_B|$ and $|Q'_I|$ vs. ψ_s in Fig. 2.7 by using (2.5.6) and (2.5.7); their sum $|Q'_C|$ is also shown. Other expressions for Q'_I are discussed in the literature.^{7,14-18}

It is convenient to divide the inversion region into three subregions: These are marked *weak*, *moderate*, and *strong* inversion in Fig. 2.7. To be consistent with the definition of inversion in Fig. 2.6, one can define the *onset* of weak inversion in

†It can be shown that more accurate expressions for Q'_C , Q'_B , and Q'_I include the term $-\phi_i$ under each square-root sign in (2.5.1), (2.5.6), and (2.5.7) (Appendix F). The error involved in omitting this term is small for the values of ψ_s assumed in inversion. In fact, the above expressions can be used successfully *even* in the upper part of depletion.

**FIGURE 2.7**

Magnitude of inversion layer charge, depletion region charge, and their sum (all per unit area) vs. surface potential.

Fig. 2.7 at $\psi_s = \phi_F$. The *upper* limit of weak inversion is defined in much of the literature at $\psi_s = 2\phi_F$. It is seen that for surface potentials less than about this value, practically all the surface charge is due to the charge in the depletion region. The corresponding inversion layer charge is too small to be shown in the scale of the figure but can nevertheless cause nonnegligible conduction when the MOS structure is part of a transistor. As ψ_s is raised above $2\phi_F$, $|Q'_I|$ starts to become significant because of the exponential in (2.5.7). For ψ_s exceeding $2\phi_F$ by a few ϕ_t , Q'_I becomes a very strong function of ψ_s . This should come as no surprise. From Fig. 2.6 we see that at $\psi_s = 2\phi_F$ the surface electron concentration is already as high as the concentration of acceptor atoms. An increase of ψ_s above $2\phi_F$ by several ϕ_t would be enough to provide very large n_{surface} because of the exponential dependence of the latter on ψ_s (2.4.12). The concentration $n(y)$ at points very close to the surface would also in-

crease drastically, and Q'_I as seen from (2.5.4) would begin "taking off." A point around which this happens can be defined as the onset of strong inversion. This point will be above $2\phi_F$ by a quantity ϕ_{Z0} , which is several ϕ_F , and will be discussed further later on.

In addition to the relative contributions of Q'_I and Q'_B to Q'_C , it is very important to consider the relative contributions to *changes* $\Delta Q'_C$, associated with changes $\Delta\psi_s$ in the surface potential. Such charge changes are provided partly by a change in the inversion layer charge and partly by a change in the depletion region charge; that is, as seen from (2.5.2), we have

$$\Delta Q'_C = \Delta Q'_I + \Delta Q'_B \quad (2.5.8)$$

Let us first consider the case where V_{GB} is low, so that ψ_s and $|Q'_C|$ are also low. Assume a change ΔV_{GB1} of V_{GB} results in a change $\Delta Q'_{C1}$ as shown in Fig. 2.7. From the figure it is clear that practically all of this change is provided by a change $\Delta Q'_{B1}$ in the depletion region charge. Furthermore, the change $\Delta\psi_{s1}$ of the surface potential, required to accommodate these charge changes, is quite significant. It is clear that what has just been described is characteristic of most of what has been marked "weak inversion" in Fig. 2.7, with the possible exception of points very close to its upper limit.

Assume now that V_{GB} is larger, causing a large ψ_s and a large $|Q'_C|$. Assume a change ΔV_{GB2} causes a change $\Delta Q'_{C2}$, as shown in Fig. 2.7. Things are now different. Practically all $\Delta Q'_C$ is provided by a change $\Delta Q'_{I2}$ of the inversion layer charge, $\Delta Q'_{B2}$ being negligible, and the surface potential change $\Delta\psi_{s2}$ required to accommodate these changes is very small. This behavior is evident throughout what has been marked "strong inversion" in Fig. 2.7.

Let us stop for a moment and summarize the most important results in our development. We have derived enough equations to characterize the MOS system in inversion; these are

1. *Potential balance.* Equation (2.3.1), repeated here for convenience, requires

$$V_{GB} = \psi_{ox} + \psi_s + \phi_{MS} \quad (2.5.9)$$

2. *Charge balance.* From (2.3.4) and (2.5.2) we have

$$Q'_G + Q'_o + Q'_I + Q'_B = 0 \quad (2.5.10)$$

3. *Relations of charges to potentials.* The final three equations relate each charge variable to the potential associated with it. In particular, the charge above the oxide is related to the potential across the oxide by (2.4.18), repeated below:

$$Q'_G = C'_{ox} \psi_{ox} \quad (2.5.11)$$

The inversion layer charge is related to the surface potential by (2.5.7), which is of the form

$$Q'_I = Q'_I(\psi_s) \quad (2.5.12)$$

Finally, the depletion region charge is related to the potential across that region by (2.5.6), which is of the form

$$Q'_B = Q'_B(\psi_s) \quad (2.5.13)$$

The five equations (2.5.9) through (2.5.13) contain six variables, three of which are potentials (V_{GB} , ψ_{ox} , ψ_s) and three of which are charges per unit area (Q'_G , Q'_I , Q'_B). Among the five equations we can eliminate four of the variables and end up with a relationship between the other two. For example, let us develop a relation between V_{GB} and ψ_s . Eliminating the other variables, we obtain (Prob. 2.3)

$$V_{GB} = V_{FB} + \psi_s - \frac{Q'_B(\psi_s) + Q'_I(\psi_s)}{C'_{ox}} \quad (2.5.14)$$

or, using (2.5.2) and (2.5.1),

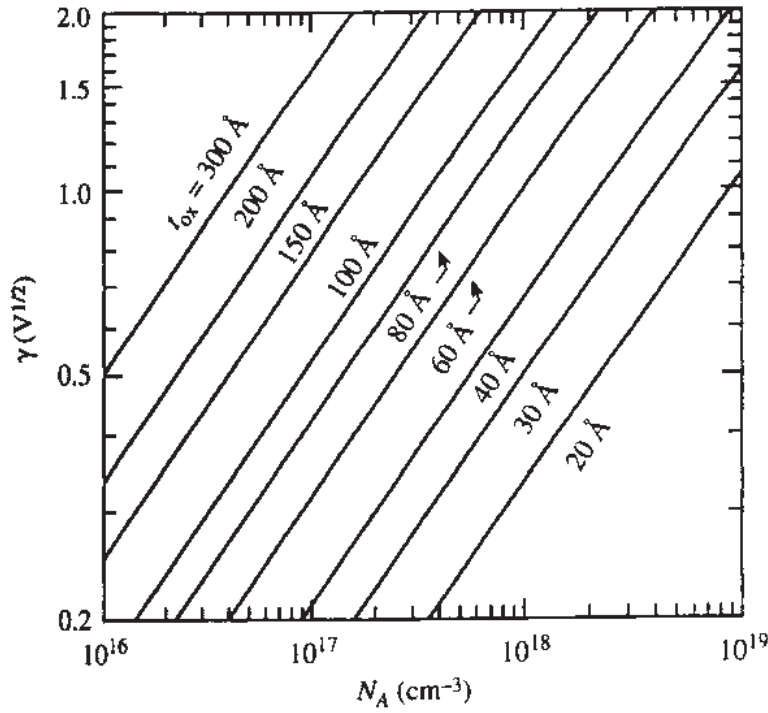
$$V_{GB} = V_{FB} + \psi_s + \gamma \sqrt{\psi_s + \phi_t} e^{(\psi_s - 2\phi_F)/\phi_t} \quad (2.5.15)$$

where

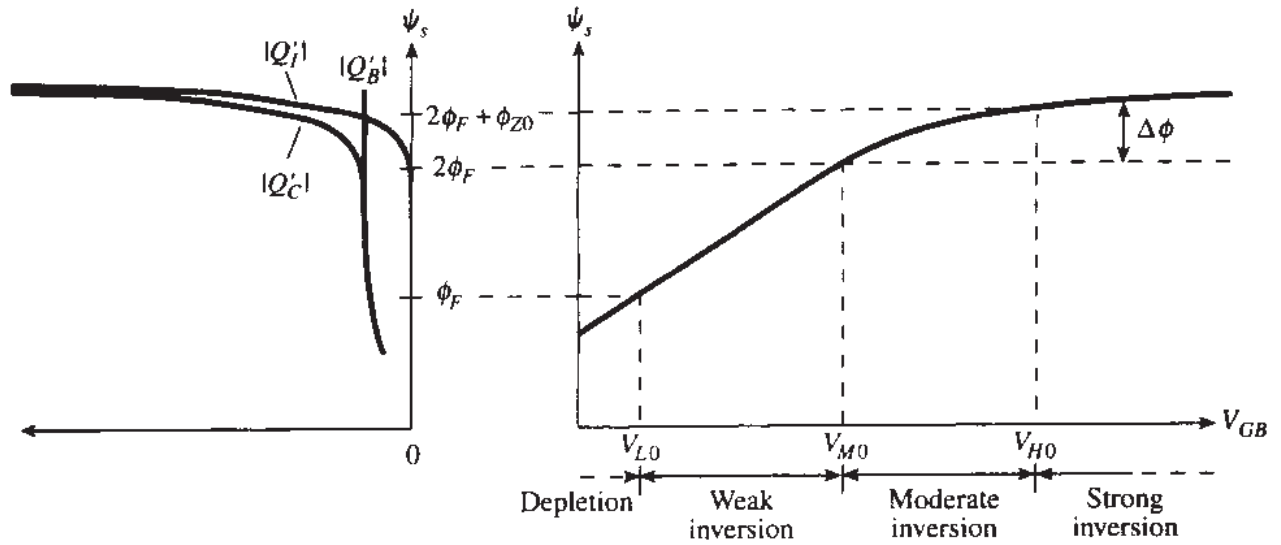
$$\gamma \equiv \frac{\sqrt{2q\epsilon_s N_A}}{C'_{ox}} \quad (2.5.16)$$

with $\sqrt{2q\epsilon_s} = 5.8 \times 10^{-16} \text{ F} \cdot \text{cm}^{-1/2} \cdot \text{V}^{1/2}$ (the units chosen here are equivalent to those given earlier and are convenient for use in the above equations). The values of γ are shown in Fig. 2.8 for a variety of process parameters. This quantity is called the *body effect coefficient*, for reasons that will become apparent in Chap. 3.

Equation (2.5.15) is plotted toward the right in Fig. 2.9, where it is shown as ψ_s vs. V_{GB} (although the equation cannot be solved explicitly for ψ_s). The plot of $|Q'_I|$ vs. ψ_s from Fig. 2.7 is repeated in the left part of Fig. 2.9, with the same ψ_s axis so that the two plots can be related. Notice that for low points on the $\psi_s(V_{GB})$ curve, significant changes $\Delta\psi_s$ are required to accommodate the charge changes demanded by ΔV_{GB} . For points high on the $\psi_s(V_{GB})$ curve, a slight $\Delta\psi_s$ is sufficient to accommodate these changes; this is in part because of the steepness of the $Q'_I(\psi_s)$ curve at such points. The weak inversion region is characterized by significant and nearly constant slope $d\psi_s/dV_{GB}$; in strong inversion, this slope drops to small values. The strong inversion region can be defined where ψ_s is practically constant (see below).

**FIGURE 2.8**

Body effect coefficient vs. substrate doping, for several values of oxide thickness.

**FIGURE 2.9**

Surface potential vs. gate-substrate voltage (right) and charges vs. surface potential (left).

In terms of V_{GB} , the onsets of weak, moderate, and strong inversion will be denoted by V_{L0} , V_{M0} , and V_{H0} , as shown in Fig. 2.9. V_{L0} and V_{M0} can be found from (2.5.15) by using in it, respectively, $\psi_s = \phi_F$ and $\psi_s = 2\phi_F$. Neglecting the resulting small exponential terms gives

$$V_{L0} = V_{FB} + \phi_F + \gamma\sqrt{\phi_F} \quad (2.5.17)$$

$$V_{M0} = V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F} \quad (2.5.18)$$

Strong inversion begins at a surface potential of $2\phi_F + \phi_{Z0}$, where ϕ_{Z0} is several ϕ_F , its value depending on substrate doping and oxide thickness. The corresponding value of V_{GB} at the onset of strong inversion, denoted by V_{H0} , is several tenths of 1 V above V_{M0} and will be discussed further in Sec. 2.5.2. More elaborate definitions¹⁹ for V_{M0} and V_{H0} are discussed in Appendix G.

It is desirable in the present development to find a relation of the form

$$Q'_I = Q'_I(V_{GB}) \quad (2.5.19)$$

since this would give directly the influence of the externally applied voltage (the "cause") on the resulting inversion layer charge (the final "effect" we are interested in, since Q'_I will result in current flow in the MOS transistor). Unfortunately, if one attempts to derive such a relation from (2.5.9) through (2.5.13), it is found that an implicit expression results; that is, Q'_I cannot be expressed in closed form as a function of V_{GB} . The solution of the resulting complicated equation has to be obtained numerically if a value for V_{GB} is given and the value for Q'_I is desired. For our purposes, instead of (2.5.19) we can consider its parametric representation, which consists of (2.5.7) and (2.5.15), i.e., the relations that were plotted in Fig. 2.9. If values are assumed for ψ_s , the corresponding Q'_I and V_{GB} can be found from these equations. Q'_I can be plotted vs. V_{GB} as in Fig. 2.10. The broken line in the figure will be discussed in the following subsection.

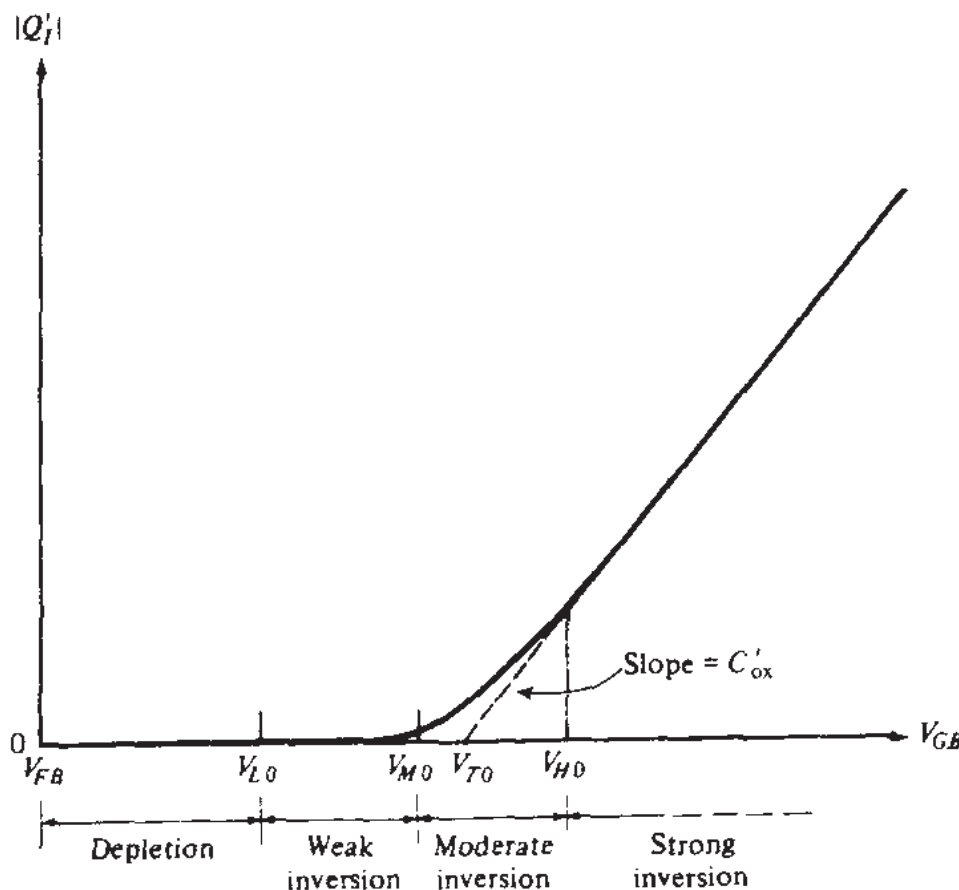


FIGURE 2.10

Magnitude of inversion layer charge per unit area vs. gate-substrate voltage.

Let us also develop an alternative expression for Q'_I for later use. If Q'_G and ψ_{ox} are eliminated among (2.5.9) to (2.5.11), and if (2.5.6) and (2.2.6) are used in the result, we obtain

$$Q'_I = -C'_{ox} (V_{GB} - V_{FB} - \psi_s - \gamma\sqrt{\psi_s}) \quad (2.5.20)$$

One can thus regard (2.5.20) and (2.5.15) as an alternative parametric representation of the relation between Q'_I and V_{GB} .

We will now consider each region of inversion (weak, moderate, and strong) separately. We will identify the dominant phenomena in each region and show certain approximations that can be made in each case. We will seek to develop explicit expressions of the form of (2.5.19), based on such approximations. This will be desirable not only for ease of computation but also because the functional form of the resulting approximate expressions will help make evident the dominant phenomena in the regions under consideration. Also, such expressions will be needed for developing simple models for the MOS transistor in subsequent chapters.

2.5.2 Strong Inversion

As seen in Fig. 2.9, in strong inversion large changes in V_{GB} result in very small changes in ψ_s . A very common assumption is that in strong inversion ψ_s is practically "pinned" to a constant value:

$$\psi_s \approx \phi_0 \quad (2.5.21)$$

The value of ϕ_0 is close to $2\phi_F + \phi_{Z0}$ in Fig. 2.9 but is difficult to define exactly. Obviously, if one is interested in operating points limited only to high V_{GB} , the value chosen for ϕ_0 would be somewhat larger than that needed for low-voltage operation.† Usually a single value is chosen for ϕ_0 as a compromise.

Historically, the most commonly used value is

$$\phi_0 = 2\phi_F \quad (2.5.22)$$

It is difficult to justify the use of this value, looking at Fig. 2.9. From that figure, it is clear that at $\psi_s = 2\phi_F$ the plot of $\psi_s(V_{GB})$ has not flattened out yet, and that in strong inversion the value of ψ_s is considerably larger than $2\phi_F$. Thus, a more appropriate value for ϕ_0 is

$$\phi_0 = 2\phi_F + \Delta\phi \quad (2.5.23)$$

†The situation is analogous to assuming a constant value for the base-emitter voltage for bipolar transistors. The value chosen depends on the context and the application and can be expected to be higher in power circuit work than in micropower circuit work.

where, for the case of uniform substrates considered so far, $\Delta\phi$ is several ϕ_i (for uniform substrates about $6\phi_i$ can be used, which is an average for a variety of process parameters and V_{GB} values). For nonuniform substrates, the analysis presented in this chapter is, in principle, not valid. Nevertheless, MOS transistor equations based on this analysis (Chap. 4) are often used for devices with nonuniform substrates and in such equations ϕ_0 appears (Chap. 5). In such cases, to achieve reasonable agreement with experimental results, $\Delta\phi$ must be modified from the estimate given above; it may or may not be close to zero, or may even be negative (Chap. 5). In any case, the historical value of $2\phi_F$ is, in general, not the best choice for ϕ_0 , so (2.5.23) should be used, with an appropriate choice for $\Delta\phi$. An approximation for $\Delta\phi$ in strong inversion is discussed in Prob. 2.5.

Equation (2.5.21) implies that the depletion region width d_B is assumed to reach a maximum value d_{Bm} , and not to increase with V_{GB} anymore. From (2.5.5) and (2.5.21),

$$d_{Bm} = \sqrt{\frac{2\epsilon_s}{qN_A}} \sqrt{\phi_0} \quad (2.5.24)$$

Similarly, the depletion region charge is assumed to have reached a maximum value given by (2.5.6) and (2.5.21):

$$Q'_B = -\sqrt{2q\epsilon_s N_A} \sqrt{\phi_0} \quad (2.5.25)$$

Using (2.5.9)–(2.5.11) and (2.5.21), we obtain

$$\boxed{Q'_I = -C'_{ox} (V_{GB} - V_{T0})} \quad (2.5.26)$$

where

$$V_{T0} = \phi_{MS} - \frac{Q'_o}{C'_{ox}} + \phi_0 - \frac{Q'_B}{C'_{ox}} \quad (2.5.27)$$

and, using (2.2.6), (2.5.25), and (2.5.16) in the above equations,

$$\boxed{V_{T0} = V_{FB} + \phi_0 + \gamma\sqrt{\phi_0}} \quad (2.5.28)$$

Equations (2.5.26) and (2.5.28) could also have been obtained directly from (2.5.20) and (2.5.21).[†] $|Q'_I|$ from (2.5.26) is shown by the broken line in Fig. 2.10, where it can be compared to the accurate $|Q'_I(V_{GB})|$ (solid line). The quantity V_{T0} is called the *extrapolated threshold voltage* of the MOS two-terminal structure, for

[†]We provide more than one way to obtain a result in an attempt to give a better “feeling” for the many new equations presented in this chapter, and also to present some intermediate results which will be needed in subsequent chapters. This practice will be repeated in the material to follow.

obvious reasons. The fact that the plot of $Q'_I(V_{GB})$ is practically a straight line in strong inversion will prove to be crucial in shaping the properties of the structure in this region.

Example 2.3. Estimate V_{T0} for the process of Example 2.2.

In Example 2.2 we found $V_{FB} = -1$ V. Using the data of that example in (2.5.16), we find $\gamma = 0.504$ V^{1/2}. Let us use $\phi_0 = 2\phi_F + 6\phi_i = 0.975$ V. Then (2.5.28) gives

$$V_{T0} = -1 \text{ V} + 0.975 \text{ V} + 0.504\sqrt{0.975} \text{ V} = 0.473 \text{ V}$$

It should be emphasized that, whereas V_{T0} is a quantity that appears in the strong inversion expression (2.5.26), the MOS structure is *not* in strong inversion at $V_{GB} = V_{T0}$; this crucial observation becomes apparent by examining Fig. 2.10.

We can now return to the discussion of the onset of strong inversion. As seen in Fig. 2.10, the transition of the exact Q'_I plot toward the straight-line behavior described by (2.5.26) is *very* gradual. No critical point can be identified that could conveniently be taken as the onset of strong inversion. Thus, a possible definition of this onset can be the minimum V_{GB} value for which (2.5.26) predicts Q'_I within an acceptable error. The value one ends up with depends, of course, on what is meant by "acceptable error." For some applications, a 10 percent error in the value of Q'_I may be tolerated. Then the onset of strong inversion can be taken relatively close to V_{T0} . If only a 2 percent error can be tolerated, one should take the onset at a somewhat larger V_{GB} value. Finally, for some applications (e.g., the evaluation of small-signal capacitances and conductances in Chap. 8), accuracy is desirable not only for the plot of Q'_I vs. V_{GB} but also for the *slope* of this plot. Then the onset of strong inversion should be defined accordingly. As is evident from the figure, at points where the slope is accurately predicted by the straight line, Q'_I itself is accurately predicted by it. Thus, definitions in terms of the slope are conservative. If a slope error of about 10 percent can be tolerated, it turns out that the onset of strong inversion V_{H0} should be taken about 0.6 V above V_{M0} at room temperature and for practical fabrication process. There are ways to define and calculate V_{M0} and V_{H0} precisely, if desired.¹⁹ These are somewhat elaborate, and are given for the interested reader in Appendix G.

The choice $\phi_0 = 2\phi_F$ (which, as we have already discussed, is in general not the best) makes V_{T0} in (2.5.28) identical to the upper limit of weak inversion, V_{M0} in (2.5.18):

$$V_{T0} = V_{M0} \quad \text{if } \phi_0 = 2\phi_F \text{ is used} \quad (2.5.29)$$

In fact, in some of the literature, the onset of strong inversion is taken as V_{M0} , and no distinction is made between the quantities V_{M0} , V_{T0} , and V_{H0} . In such treatments, all three quantities are taken to be the same point, called "threshold," and assumed to be given by (2.5.18). No moderated inversion is defined in such cases, and the strong-inversion region is taken to be adjacent to weak inversion. This practice originates in the early days of MOS work, when large V_{GB} values were common. If, for example, the actual values of V_{M0} , V_{T0} , and V_{H0} happened to be 1.0, 1.2, and 1.6 V, respectively, and the V_{GB} values of interest were 20 to 30 V, then using 1.0 or 1.6 V in lieu of V_{T0} in (2.5.26) would not result in excessive error. Today, however,

with V_{GB} values limited to about 3 V for many practical cases, and with a constant drive to even lower voltages, a careful distinction between the three voltages is necessary.

2.5.3 Weak Inversion

In weak inversion $Q'_I(V_{GB})$ will be shown to assume a simple form, albeit very different from that in strong inversion. To start, consider (2.5.7) and define:

$$\xi = \phi_i e^{(\psi_s - 2\phi_F)/\phi_i} \quad (2.5.30)$$

Consider the term $\sqrt{\psi_s + \xi}$ in (2.5.7). In weak inversion, ψ_s is smaller than $2\phi_F$, as seen in Fig. 2.9. Thus $\xi \ll \psi_s$. The function $\sqrt{\psi_s + \xi}$ can be approximated by the first two terms of its Taylor expansion around $\xi = 0$:

$$\sqrt{\psi_s + \xi} = \sqrt{\psi_s} + \frac{1}{2\sqrt{\psi_s}} \xi \quad (2.5.31)$$

Using (2.5.30) and (2.5.31) in (2.5.7) we obtain

$$Q'_I \approx -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\psi_s}} \phi_i e^{(\psi_s - 2\phi_F)/\phi_i} \quad (2.5.32)$$

To obtain a relation between Q'_I and the external bias V_{GB} , we need to relate ψ_s in the above equation to V_{GB} . This can be done very easily by noting that the general $V_{GB}(\psi_s)$ relation, (2.5.14), can be simplified to

$$V_{GB} \approx V_{FB} + \psi_s - \frac{Q'_B(\psi_s)}{C'_{ox}} \quad (2.5.33a)$$

because in weak inversion $|Q'_I| \ll |Q'_B|$. From (2.5.6) and (2.5.16), this becomes

$$V_{GB} \approx V_{FB} + \psi_s + \gamma\sqrt{\psi_s} \quad (2.5.33b)$$

which could also have been obtained from (2.5.15) by noting that the exponential in the equation is negligible. Solving (2.5.33b), we have

$$\psi_s \approx \psi_{sa} \quad (2.5.34)$$

with

$$\psi_{sa} = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right)^2 \quad (2.5.35)$$

It is seen from the above development that ψ_{sa} is the value of the surface potential that develops when the inversion layer charge is negligible. In Fig. 2.11, we re-

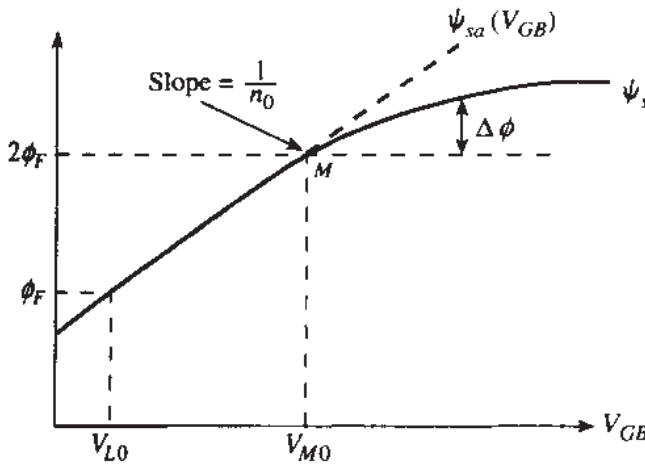


FIGURE 2.11

Surface potential and potential ψ_{sa} vs. gate-substrate voltage.

peat the $\psi_s(V_{GB})$ plot of Fig. 2.9, and we show $\psi_{sa}(V_{GB})$ for comparison. It is seen that the agreement in weak inversion is very good, except perhaps at points very close to the upper limit of the region. We can thus use $\psi_s \approx \psi_{sa}$ in (2.5.32) to obtain:

$$Q'_I = -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\psi_{sa}(V_{GB})}} \phi_i e^{[\psi_{sa}(V_{GB}) - 2\phi_F]/\phi_i} \quad (2.5.36)$$

This equation, with $\psi_{sa}(V_{GB})$ given by (2.5.35), is an *explicit* expression of Q'_I in terms of V_{GB} .

It can be seen in Fig. 2.11 that the slope of ψ_{sa} (and thus of ψ_s in weak inversion) vs. V_{GB} is almost constant. The *inverse* of this slope is often denoted by n . Thus

$$n \equiv \left(\frac{d\psi_{sa}}{dV_{GB}} \right)^{-1} \quad (2.5.37)$$

The value of n can be found in terms of V_{GB} by using (2.5.35) in the above definition, which gives:

$$n = 1 + \frac{\gamma}{2\sqrt{\psi_{sa}(V_{GB})}} \quad (2.5.38)$$

As seen, n is larger than unity.† This fact could have been anticipated intuitively, since changes in V_{GB} appear partly as changes in ψ_{ox} and partly as changes in ψ_s , and thus $\Delta\psi_s < \Delta V_{GB}$. Since $\psi_s \approx \psi_{sa}$, this means $\Delta\psi_{sa}/\Delta V_{GB} < 1$, which implies $n > 1$ from (2.5.37). Typically, n is between 1 and 1.5.

†If the interface trap charge (Sec. 2.2) varies with the surface potential, an extra term will be needed in this equation. This is shown at the end of Sec. 2.6.

The relation of Q'_I to V_{GB} , given by (2.5.36) and (2.5.35), is rather complicated. A simpler approximate form can be obtained.²⁰⁻²² First, note that when V_{GB} is varied, the variation of $\sqrt{\psi_{sa}}$ in (2.5.36) is negligible when compared to the drastic variation of the exponential term in that equation; thus we assume that $\sqrt{\psi_{sa}}$ is practically fixed and we can replace that term by $\sqrt{\phi_x}$, where ϕ_x is the value of the surface potential at any convenient point in weak inversion. Simple formulas result if we choose^{20,21} $\phi_x = 2\phi_F$ †:

$$Q'_I \approx -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{2\phi_F}} \phi_t e^{[\psi_{sa}(V_{GB}) - 2\phi_F]/\phi_t} \quad (2.5.39)$$

This equation can be further simplified by taking advantage of the almost constant slope of ψ_{sa} with V_{GB} , a fact already discussed above. At $\psi_{sa} = 2\phi_F$ the value of n [which is the inverse of the slope $d\psi_{sa}/dV_{GB}$, as seen from (2.5.37)] is, from (2.5.38):

$$n_0 = n|_{\psi_{sa}=2\phi_F} = 1 + \frac{\gamma}{2\sqrt{2\phi_F}} \quad (2.5.40)$$

Thus, taking changes *below* point M in Fig. 2.11, we can write

$$\psi_{sa} - 2\phi_F = \frac{1}{n_0} (V_{GB} - V_{M0}) \quad (2.5.41)$$

Using this in (2.5.39), we obtain

$$Q'_I \approx Q'_{M0} e^{(V_{GB} - V_{M0})/(n_0 \phi_t)} \quad (2.5.42)$$

with

$$Q'_{M0} = -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{2\phi_F}} \phi_t \quad (2.5.43)$$

Q'_{M0} represents the value of Q'_I at the upper limit of weak inversion ($V_{GB} = V_{M0}$).‡ Here one may worry that, at that point, the slope of $\psi_s(V_{GB})$ is not representative of the slope lower in the weak inversion region (Fig. 2.11). However, note that n_0 in the above development was *not* defined in terms of $\psi_s(V_{GB})$, but rather in terms of $\psi_{sa}(V_{GB})$, the slope of which does *not* change drastically at $V_{GB} = V_{M0}$, as seen in Fig. 2.11.

†Obviously, a more appropriate point would be $\phi_x = 1.5\phi_F$, which lies at the middle of weak inversion.²² However, choosing $\phi_x = 2\phi_F$ results in simpler results, and is consistent with the level of approximation we are seeking at this point.

‡As indicated in (2.5.29), V_{T0} coincides with this limit if ϕ_0 is taken equal to $2\phi_F$. The appropriateness of this choice has been questioned in Sec. 2.5.2.

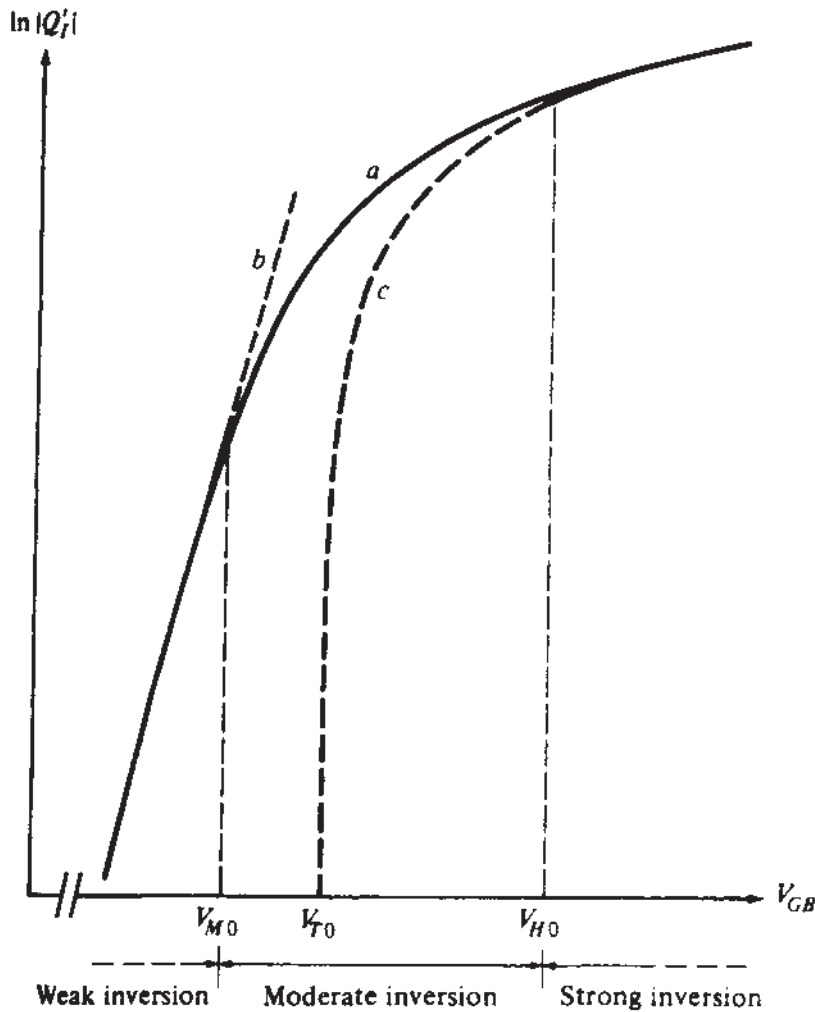


FIGURE 2.12

Logarithm of inversion layer charge magnitude per unit area vs. gate-substrate voltage. (a) Equations (2.5.7) and (2.5.15); (b) (2.5.42), (c) (2.5.26).

In Fig. 2.12, we have plotted $\ln |Q'_I|$ vs. V_{GB} , using (2.5.42) (line *b*).† The solid line represents the accurate equations (2.5.7) and (2.5.15), and describes the structure correctly for all regions of inversion. Over the weak-inversion region, the agreement is seen to be very good except at points very close to the upper limit of the region. In weak inversion, the plot of $\ln |Q'_I|$ is seen to be practically a straight line, verifying that $|Q'_I|$ is nearly exponential with respect to V_{GB} .

As has been seen, the development of (2.5.42) involved a number of approximations. Thus, although this equation and Fig. 2.12 suggest that $Q'_I(V_{GB})$ is exponential in weak inversion, it should be kept in mind that this is not *exactly* so. Indeed,

†In showing the logarithm of various quantities in this book, we assume that these quantities have been normalized to their units, so that they are rendered dimensionless before taking their logarithm. For simplicity, we do not define new symbols for these dimensionless quantities.

more accurate equations, such as (2.5.7) with (2.5.15), or even (2.5.36), predict minute deviations from “exponentiality.” Although such deviations are too small to be visible in the plot of $\ln |Q'_I|$ vs. V_{GB} in Fig. 2.12, they can make a difference in a plot of the *slope* of $\ln |Q'_I|$ vs. V_{GB} . This can affect the evaluation of small-signal quantities (Sec. 2.6 and Chap. 8). In order to take this effect into account, the quantity n_0 is sometimes allowed to be a weak function of V_{GB} as in (2.5.38).

To conclude this section, we note that from (2.5.38) and (2.5.16) we have

$$(n - 1)C'_{ox} = \frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\psi_{sa}(V_{GB})}} \quad (2.5.44)$$

and from (2.5.40) and (2.5.16) we have:

$$(n_0 - 1)C'_{ox} = \frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{2\phi_F}} \quad (2.5.45)$$

The quantities on the right-hand side of these equations appear in (2.5.36) and (2.5.39). Thus, in some of the literature these quantities are replaced in such equations by $(n - 1)C'_{ox}$ and $(n_0 - 1)C'_{ox}$, respectively. We did not follow this practice, since such expressions do not make apparent the fact that these quantities are, in fact, independent of C'_{ox} (unless one happens to remember that $n - 1$ and $n_0 - 1$ are inversely proportional to C'_{ox}).

2.5.4 Moderate Inversion

In moderate inversion, none of the simplifications discussed above is valid. The plot of $Q'_I(V_{GB})$ here is neither a straight line nor an exponential. If accurate results are needed in this region, the complete equations (2.5.7) and (2.5.15) can be used. However, this will require a numerical solution for a given V_{GB} value, as (2.5.15) is an implicit equation in ψ_s . Efforts have been made to produce explicit expressions for ψ_s and for the charges in terms of V_{GB} directly.^{14-18,7,23} Such efforts involve several approximations.

For the interested reader, we now sketch some representative conclusions of such treatments.[†] It is clear from Fig. 2.9 that the form of the relation we have used for ψ_s in strong inversion in Sec. 2.5.2, namely

$$\psi_s = 2\phi_F + \Delta\phi \quad (2.5.46)$$

can be used in moderate inversion, too, but here $\Delta\phi$ must clearly be allowed to be a function of V_{GB} . It has been suggested that an approximate expression for $\Delta\phi$ is¹⁸

$$\Delta\phi = \frac{2\phi_F}{n} \ln \left(1 + \frac{\psi_{sa} - 2\phi_F}{2\phi_F} \right) \quad (2.5.47)$$

[†]Material set in smaller type can be skipped without loss of continuity.

where n and ψ_{sa} are both functions of V_{GB} and are given by (2.5.38) and (2.5.35), respectively. Note that, as is easily confirmed by Fig. 2.9, as V_{GB} is reduced $\Delta\phi$ decreases; when V_{GB} becomes so low that $\psi_{sa} = 2\phi_F$ the argument of the logarithm in the above equation becomes 1, and $\Delta\phi$ becomes 0; thus $\psi_s = 2\phi_F$ at that point, which is precisely what it should have been at the boundary between moderate and weak inversion. The results of Sec. 2.5.3 take over for lower values of V_{GB} (the above equation should *not* be used in weak inversion, as it would provide meaningless results in that region).

To provide continuity from moderate to strong inversion, the above $\Delta\phi$ can be used in strong inversion, too. In fact, doing so will slightly improve the accuracy there, in comparison to using a constant $\Delta\phi$.

For the inversion layer charge in moderate inversion, a similar approach has been suggested¹⁸:

$$Q'_I = -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\psi_{sa}}} \phi_i - nC'_{ox}(\psi_{sa} - 2\phi_F - \Delta\phi) \quad (2.5.48)$$

Again, when V_{GB} is reduced to the point that $\psi_{sa} = 2\phi_F$, the second term in the above equation becomes 0, and Q'_I reduces to the value predicted at that point by the weak inversion expression, (2.5.36). The latter equation [or (2.5.42)] must take over for lower V_{GB} values, as (2.5.48) is not valid in weak inversion. As was the case for ψ_s , to provide continuity from moderate to strong inversion one can use (2.5.48) in strong inversion also, where it has been claimed to provide good accuracy.¹⁸

As has already been noted, in many treatments no moderate-inversion region is defined. Sometimes this region is considered the bottom part of strong inversion. In some treatments a point is defined, falling somewhere in our moderate-inversion region, and $Q'_I(V_{GB})$ is assumed to be exponential directly below and a straight line directly above that point. Such models can lead to large errors. This can be seen in Fig. 2.12, where Q'_I as calculated from (2.5.7) and (2.5.15) is compared to the approximate equations (2.5.42) and (2.5.26). *Neither* of these two equations provides satisfactory accuracy in moderate inversion.

2.6 SMALL-SIGNAL CAPACITANCE

If V_{GB} is increased by a small amount ΔV_{GB} in Fig. 2.4a, a positive charge† $\Delta Q'_G$ will flow into the gate terminal. For overall charge neutrality, a charge of equal value must flow out of the substrate terminal or, equivalently, a charge of value $-\Delta Q'_G$ must flow *into* the substrate terminal. An incremental (small-signal) capacitance per unit area, C'_{gb} , can thus be defined to relate charge changes to voltage changes. This is illustrated in Fig. 2.13. We define

$$C'_{gb} \equiv \frac{dQ'_G}{dV_{GB}} \quad (2.6.1)$$

†We use the term "charge" rather than "charge per unit area" for simplicity. The fact that we are talking about charges *per unit area* will be clear from the primes used in the symbols for these quantities.



FIGURE 2.13

Small-signal equivalent circuit for a two-terminal MOS structure driven by a voltage source.

The charge $-\Delta Q'_G$ flowing into the substrate goes to change the substrate charge Q'_C by an amount $\Delta Q'_C$:

$$\Delta Q'_C = -\Delta Q'_G \quad (2.6.2)$$

The gate-to-substrate voltage change will be distributed partly across the oxide (as $\Delta\psi_{ox}$) and partly across the semiconductor (as a change $\Delta\psi_s$ in the surface potential). Equation (2.3.2) is repeated here for convenience:

$$\Delta V_{GB} = \Delta\psi_{ox} + \Delta\psi_s \quad (2.6.3)$$

Taking $1/C'_{gb}$ from (2.6.1) and using (2.6.3) in the result (with small changes replaced by differentials), we have

$$\frac{1}{C'_{gb}} = \frac{d\psi_{ox}}{dQ'_G} + \frac{d\psi_s}{dQ'_G} \quad (2.6.4)$$

The above equation can be written, by using (2.6.2), as

$$\frac{1}{C'_{gb}} = \frac{1}{dQ'_G/d\psi_{ox}} + \frac{1}{-dQ'_C/d\psi_s} \quad (2.6.5)$$

To interpret this equation, note first that from (2.5.11) we have

$$\frac{dQ'_G}{d\psi_{ox}} = C'_{ox} \quad (2.6.6)$$

The quantity $-dQ'_C/d\psi_s$ in (2.6.5) can be interpreted with the help of Fig. 2.4a as follows. If the potential across the semiconductor is changed by $\Delta\psi_s$, the charge in that region will change by $\Delta Q'_C$. This additional charge must enter the region through its *bottom*, coming from the substrate terminal. If now a capacitor had across it a voltage $\Delta\psi_s$, and the value of its capacitance were adjusted so that a charge $\Delta Q'_C$ had to flow into its *bottom* plate, the situation illustrated in Fig. 2.14 would result. From this figure we clearly have

$$C'_c \equiv -\frac{dQ'_C}{d\psi_s} \quad (2.6.7)$$

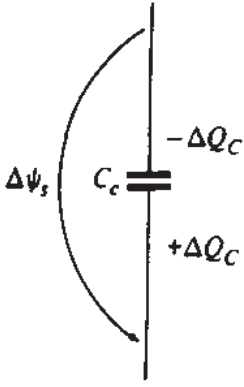


FIGURE 2.14
Illustrating the concept of semiconductor charge region small-signal capacitance.

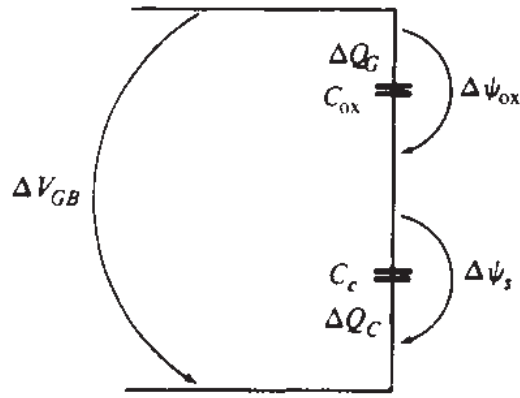


FIGURE 2.15
Circuit representation of (2.6.8).

The denominator of the last fraction in (2.6.5), therefore, can be interpreted as a small-signal capacitance corresponding to the semiconductor charge region. It relates the changes of the potential across that region to the corresponding changes in its charge.

Using now (2.6.6) and (2.6.7) in (2.6.5), we get

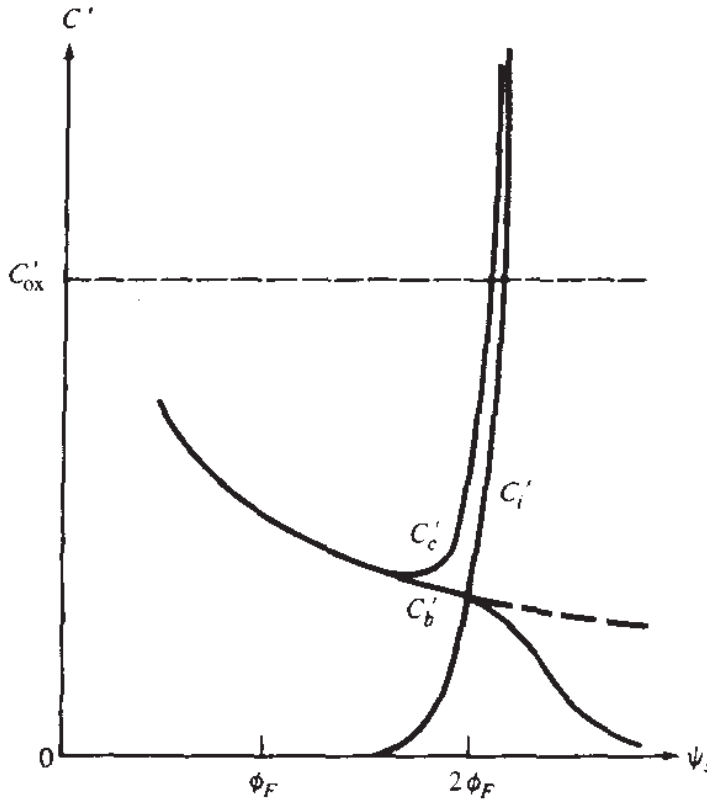
$$\frac{1}{C'_{gb}} = \frac{1}{C'_{ox}} + \frac{1}{C'_c} \quad (2.6.8)$$

Therefore the small-signal capacitance C'_{gb} is the same as that exhibited by two capacitors of values C'_{ox} and C'_c , connected in series as in Fig. 2.15. C'_{gb} can be evaluated from (2.6.8), (2.6.7), and (2.4.17) (the general expression for C'_c is given in Appendix F).

Let us consider the special case of accumulation. If ψ_s is negative and its absolute value is at least a few ϕ_i , Q'_C in (2.4.17) is a very sensitive function of ψ_s , and C'_c from (2.6.7) is very large. Then the total capacitance C'_{gb} is reduced to approximately C'_{ox} , as seen from (2.6.8). Intuitively, this should make sense. In accumulation there is an abundance of holes which can provide a conducting path from the substrate bottom through the semiconductor to the surface. For V_{GB} sufficiently smaller than V_{FB} , the resulting negative surface potential attracts huge numbers of holes immediately below the oxide, in a sense forming the bottom "plate" of the oxide capacitor. As a result, the total incremental capacitance seen between the two terminals of the MOS junction is basically that of the oxide, C'_{ox} .

Consider now the case of ψ_s positive and larger than a few ϕ_i (say, larger than $3\phi_i$; this includes the upper part of depletion, as well as the inversion region). Now there are practically no holes at the surface, and Q'_C is given by (2.5.1). Using this equation in (2.6.7), we obtain

$$C'_c = \sqrt{2q\epsilon_s N_A} \frac{1 + e^{(\psi_s - 2\phi_F)/\phi_i}}{2\sqrt{\psi_s + \phi_i e^{(\psi_s - 2\phi_F)/\phi_i}}}, \quad \psi_s > 3\phi_i \quad (2.6.9)$$

**FIGURE 2.16**

Small-signal capacitances per unit area vs. surface potential. C'_i is the inversion layer capacitance; C'_b is the depletion region capacitance (solid line is exact; broken line is as predicted by the charge sheet model); C'_c is the semiconductor charge region capacitance; C'_{ox} is the oxide capacitance.

A plot of C'_c vs. ψ_s is shown in Fig. 2.16. The rest of the plots in the figure will be discussed later in this section.

It is interesting to consider the individual contributions of the depletion region and inversion layer charges to C'_c . Equation (2.5.8) is repeated below:

$$\Delta Q'_C = \Delta Q'_B + \Delta Q'_I \quad (2.6.10)$$

Using this equation in (2.6.7) we obtain

$$C'_c = \frac{-dQ'_B}{d\psi_s} + \frac{-dQ'_I}{d\psi_s} \quad (2.6.11)$$

We have then separated the total semiconductor capacitance C'_c into two components, one owing to the depletion region charge and one owing to the inversion layer charge. With a reasoning analogous to that preceding (2.6.7), we can define a depletion region incremental capacitance per unit area:

$$C'_b \equiv - \frac{dQ'_B}{d\psi_s} \quad (2.6.12)$$

This capacitance relates changes of the potential across the depletion region to the associated changes of the charge in it.

Let us also define a capacitance per unit area associated with the inversion layer. This capacitance should relate changes in the charge of that layer to the associated potential changes. In analogy with (2.6.12),

$$C'_i \equiv - \frac{dQ'_i}{d\psi_s} \quad (2.6.13)$$

From the above definitions it is seen that C'_b and C'_i are the slopes of plots $|Q'_B|$ vs. ψ_s and $|Q'_I|$ vs. ψ_s , respectively. However, to evaluate C'_b and C'_i one would have to use *very* accurate expressions for Q'_B and Q'_I , since even a small error in predicting a certain function can cause a severe error in predicting its derivative. For example, consider $|Q'_B|$ as predicted by the charge sheet model in (2.5.6) and plotted in Fig. 2.7. The corresponding C'_b is shown by the broken line in Fig. 2.16. A more exact model would be based on the general analysis outlined in Sec. 2.4, which allows for the spreading of the inversion layer into the depletion region and for the presence of holes there. Such a model gives a $|Q'_B|$ plot which, in moderate and strong inversion, is flatter than that shown in Fig. 2.7, becoming practically horizontal in very strong inversion. The corresponding error in the approximate $|Q'_B|$ is very small; yet, the slopes of the exact and approximate plots are *very* different (see solid line for C'_b in Fig. 2.16). The resulting exact C'_b and C'_i are given by (Appendix F)

$$C'_b = \sqrt{2q\epsilon_s N_A} \frac{1}{2\sqrt{\psi_s + \phi_t} e^{(\psi_s - 2\phi_F)/\phi_t}} \quad (2.6.14)$$

$$C'_i = \sqrt{2q\epsilon_s N_A} \frac{e^{(\psi_s - 2\phi_F)/\phi_t}}{2\sqrt{\psi_s + \phi_t} e^{(\psi_s - 2\phi_F)/\phi_t}} \quad (2.6.15)$$

It is easy to see that, using Q'_B as predicted by the charge sheet model in (2.5.6), we would have obtained (2.6.14) without the exponential term; that term, however, becomes dominant in strong inversion. Note also that using the accurate formulas above we obtain $C'_i = C'_b$ at $\psi_s = 2\phi_F$.

Using (2.6.12) and (2.6.13) in (2.6.11) we obtain

$$C'_c = C'_b + C'_i \quad (2.6.16)$$

and (2.6.8) becomes

$$\frac{1}{C'_{gb}} = \frac{1}{C'_{ox}} + \frac{1}{C'_b + C'_i} \quad (2.6.17)$$

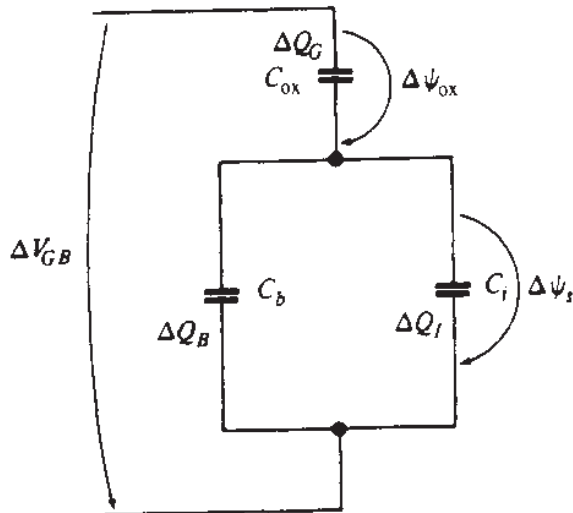


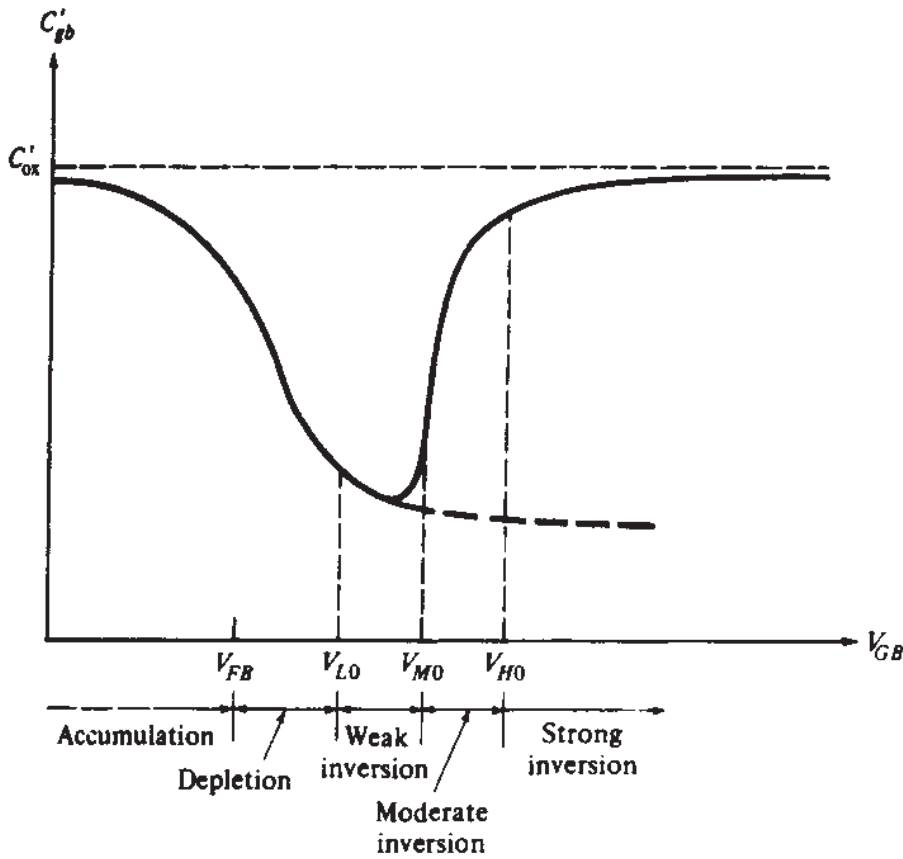
FIGURE 2.17

Small-signal equivalent circuit for the two-terminal MOS structure, showing individual capacitances, and assuming zero interface trap charge and very low frequency operation.

which can be represented by the circuit of Fig. 2.17. It should be kept in mind that this is a *small-signal equivalent* circuit, relating small *changes* of potentials and charges around a bias point. It does *not* relate total values of potentials and charges.

Our ultimate goal in this development is to plot the total capacitance seen externally (C'_{gb}) vs. the total externally applied bias (V_{GB}). This can be done as follows. For a given ψ_s , C'_c is determined; then C'_{gb} is found from (2.6.8). The result is shown by the solid curve in Fig. 2.18. We see that deep in accumulation, C'_{gb} approaches C'_{ox} , as we intuitively predicted above. For V_{GB} in the weak-inversion region (except for points close to the upper limit of the region), the inversion layer capacitance is negligible, as we can see from Fig. 2.16. From (2.6.17), then, C'_{gb} is basically the series combination of C'_{ox} and C'_b . As V_{GB} is increased, C'_b becomes smaller, as seen from Fig. 2.16. Therefore, the series combination of C'_{ox} and C'_b also decreases, as seen in Fig. 2.18. Above weak inversion, C'_i becomes significant and drastically increases if V_{GB} is raised further (Fig. 2.16). This capacitance is in parallel with C'_b . The last fraction in (2.6.17) decreases drastically, and C'_{gb} approaches C'_{ox} . Physically, an abundance of electrons exists at high V_{GB} immediately below the oxide and provides the bottom "plate" of the oxide capacitor, just as an abundance of holes provided that plate in the case of accumulation.

The above discussion has been in terms of "static" changes, i.e., it is assumed that after V_{GB} is changed by a small amount ΔV_{GB} , it remains fixed at its new value. We then wait long enough for a new equilibrium to be reached, and we record the changes in the various potentials and charges. If ΔV_{GB} is a small-signal sinusoidal voltage, the steady-state charge changes will also be sinusoidal. They will correspond to equilibrium values *if* the frequency is low enough (e.g., 1 Hz). However, if the frequency is high (e.g., 100 kHz), things will be different, as shown by the broken line in Fig. 2.18. The inversion layer charge now cannot keep up with the fast-changing ΔV_{GB} , and the required charge changes must be provided by covering or uncovering acceptor atoms at the bottom of the depletion region, just as in the case of depletion operation. The reason the inversion layer charge cannot follow fast enough is that it is in a sense isolated from the outside world by the oxide on top and the depletion region below. Therefore, the electron concentration there can be changed only by the

**FIGURE 2.18**

Total gate-substrate capacitance per unit area vs. gate-substrate bias. Solid line: "static" behavior; broken line: high-frequency behavior.

mechanisms of thermal generation and recombination, which in this case are very slow (no external irradiation is assumed). If, instead, communication with the outside world were possible, in the sense that inversion layer charge could be provided or removed externally, then the behavior exhibited by the solid curve in strong inversion would persist up to much higher frequencies. This communication with the outside world is provided by the source and drain regions in a MOS transistor, as discussed in subsequent chapters.

The incremental capacitances defined in this section can be used to provide convenient expressions for the slopes of various plots in Sec. 2.5. Thus, using (2.5.14), (2.6.12), and (2.6.13) we have

$$\frac{d\psi_s}{dV_{GB}} = \frac{C'_{ox}}{C'_{ox} + C'_b + C'_i} \quad (2.6.18)$$

which could also have been obtained from the equivalent circuit in Fig. 2.17. This relation gives simply the slope of the ψ_s vs. V_{GB} plot of Fig. 2.9. Similarly, the slope of the $|Q'_I|$ vs. V_{GB} plot in Fig. 2.10 can be found as follows:

$$\frac{d|Q'_I|}{dV_{GB}} = \frac{d|Q'_I|}{d\psi_s} \frac{d\psi_s}{dV_{GB}} \quad (2.6.19)$$

and, using (2.6.13) and (2.6.18),

$$\frac{d|Q'_I|}{dV_{GB}} = \frac{C'_{ox} C'_i}{C'_{ox} + C'_b + C'_i} \quad (2.6.20)$$

Finally, the slope of the $\ln |Q'_I|$ vs. V_{GB} plot in Fig. 2.12 can be found by using (2.6.20):

$$\frac{d \ln |Q'_I|}{dV_{GB}} = \frac{C'_{ox}}{C'_{ox} + C'_b + C'_i} \frac{C'_i}{|Q'_I|} \quad (2.6.21)$$

Deep in weak inversion C'_i can be neglected in (2.6.18) (Fig. 2.16). The inverse of the quantity $d\psi_s/dV_{GB}$ in this region is practically equal to the quantity denoted by n in Sec. 2.5 [see (2.5.37)]; then, from (2.6.18):

$$n = 1 + \frac{C'_b}{C'_{ox}} \quad (2.6.22)$$

which can be easily shown to reduce to (2.5.38).

It should be noted here that throughout our discussion so far we have assumed that the equivalent interface charge Q'_o is fixed and independent of voltage. This may not be accurate in some devices (especially those fabricated by older techniques) in which a significant density of interface traps may exist (Sec. 2.2). These traps, located at the oxide-silicon interface, can exchange carriers with the silicon. The charge trapped in them depends on the value of the surface potential ψ_s . Thus, let Q'_{it} represent the fraction of Q'_o associated with interface traps. Then, in going from (2.3.4) to (2.3.5), one should include the term $\Delta Q'_{it}$ in the latter. If this term is also included in (2.6.2) and in the development following that equation, it can be easily seen that we can take the variation of Q'_{it} with ψ_s into account in the same way that we took into account the variation of Q'_B and Q'_I with ψ_s ; that is, we can define an incremental capacitance corresponding to the interface traps in analogy with (2.6.12) and (2.6.13). This capacitance will then be

$$C'_{it} \equiv - \frac{dQ'_{it}}{d\psi_s} \quad (2.6.23)$$

C'_{it} will appear in parallel with C'_b and C'_i in Fig. 2.17. If it is significant, the various formulas should be modified to include it. For example, (2.6.22) would be modified as follows:²⁰

$$n = 1 + \frac{C'_b + C'_{it}}{C'_{ox}} \quad (2.6.24)$$

Measurements performed on MOS transistors have shown that, with well-developed fabrication processes, C'_{it} is often much smaller than C'_b and may be neglected.²¹

TABLE 2.1
Regions of inversion and properties

| | Weak inversion | Moderate inversion | Strong inversion |
|--|---|---|--|
| Definition in terms of surface potential ψ_s | $\phi_F \leq \psi_s < 2\phi_F$ | $2\phi_F \leq \psi_s < 2\phi_F + \phi_{Z0}$ | $2\phi_F + \phi_{Z0} \leq \psi_s$ |
| Definition in terms of gate-substrate voltage V_{GB} | $V_{L0} \leq V_{GB} < V_{M0}$ | $V_{M0} \leq V_{GB} < V_{H0}$ | $V_{H0} \leq V_{GB}$ |
| $\frac{ Q_i }{ Q_b }$ | $\ll 1$ | Varies | $\gg 1$ deep in strong inversion; not necessarily so near the bottom of the region |
| $\frac{C_i}{C_b}$ | $\ll 1$ deep in weak inversion; not necessarily so near the top of the region | Varies | $\gg 1$ |
| $\frac{d\psi_s}{dV_{GB}}$ | Approximately constant; attains its maximum value in this region | Varies | Small |
| Dependence of Q_i on V_{GB} | Approximately exponential | — | Approximately first-degree polynomial |
| $\frac{d \ln Q_i }{d\psi_s}$ | $\frac{1}{\phi_i}$ | Varies | $\frac{1}{2\phi_i}$ |

2.7 SUMMARY OF PROPERTIES OF THE REGIONS OF INVERSION

From the relations given in this chapter, and from the associated discussions, we can summarize certain important properties for each region of inversion as shown in Table 2.1. Most of these properties have already been discussed; the last one is considered in Prob. 2.17.

REFERENCES

1. E. H. Nicollian and J. R. Brews, *MOS Physics and Technology*, John Wiley, New York, 1982.
2. A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley, New York, 1967.
3. S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.
4. R. F. Pierret, *Field Effect Devices*, Addison-Wesley, Reading, 1983.
5. H. C. de Graaff and F. M. Klaassen, *Compact Transistor Modelling for Circuit Design*, Springer-Verlag, Vienna, 1990.
6. N. Arora, *MOSFET Models for VLSI Circuit Simulation-Theory and Practice*, Springer-Verlag, Vienna, 1993.

7. K. Lee, M. Shur, T. A. Fjeldly, and T. Ytterdal, *Semiconductor Device Modeling for VLSI*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
8. I. R. M. Mansour, "On the modeling of MOS devices," *Proceedings of the Third International Symposium on Network Theory*, Yugoslavia, pp. 705–713, 1975.
9. I. R. M. Mansour, "Improved modeling of MOS devices," *Proceedings of the European Conference on Circuit Theory and Design*, Italy, 1976.
10. G. Baccarani, M. Rudan, and G. Spadini, "Analytical i.g.f.e.t. model including drift and diffusion currents," *IEEE Journal of Solid-State and Electron Devices*, vol. 2, pp. 62–68, March 1978.
11. J. R. Brews, "A charge-sheet model of the MOSFET," *Solid-State Electronics*, vol. 21, pp. 345–355, 1978.
12. F. Van de Wiele, "A long-channel MOSFET model," *Solid-State Electronics*, vol. 22, pp. 991–997, 1979.
13. J. R. Brews, "Physics of the MOS transistor," chap. 1 in *Silicon Integrated Circuits, Part A*, D. Kahng (editor), Applied Solid State Science Series, Academic Press, New York, 1981.
14. L. L. Lewyn and J. D. Meindl, "An IGFET inversion charge model for VLSI systems," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 434–440, February 1985.
15. M. A. Maher and C. A. Mead, "A physical charge-controlled model for MOS transistors," pp. 211–229, from P. Lesleben (editor), *Advanced Research VLSI*. The MIT Press, Cambridge, Mass., 1987.
16. C.-K. Park, C.-Y. Lee, K. Lee, B.-J. Moon, Y. H. Byun, and M. Shur, "A unified current-voltage model for long-channel nMOSFET's," *IEEE Transactions on Electron Devices*, vol. 38, pp. 399–406, February 1991.
17. B. Iñiguez and E. G. Moreno, "Explicit C_{∞} -continuous and general model for nMOSFETs," *Electronics Letters*, vol. 29, pp. 1036–1037, 27 May 1993.
18. A. I. A. Cunha, M. C. Schneider, and C. G. Montoro, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electronics*, vol. 38, pp. 1945–1952, November 1995.
19. Y. Tsididis, "Moderate inversion in MOS devices," *Solid-State Electronics*, vol. 25, pp. 1099–1104, 1982; see also Erratum, *ibid.*, vol. 26, p. 823, 1983.
20. R. M. Swanson and J. D. Meindl, "Ion-implanted complementary MOS transistors in low-voltage circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-7, pp. 146–153, April 1972.
21. E. Vittoz and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operation," *IEEE Journal of Solid-State Circuits*, vol. SC-12, pp. 224–231, June 1977.
22. R. J. Overstraeten, G. J. Declercq, and P. A. Nuls, "Theory of the MOS transistor in weak inversion—new method to determine the number of surface states," *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 282–288, May 1975.
23. V. Altschul and Y. S.-Diamand, "Modeling of the MOSFET inversion charge and drain current in moderate inversion," *IEEE Transactions on Electron Devices*, vol. 37, pp. 1909–1915, August 1990.

PROBLEMS

- 2.1. Show that, independently of how many different materials are used in the external path in Fig. 2.2c, the value of V_{GB} needed to make the charges disappear is given by (2.2.2).
- 2.2. Calculate the flat-band voltage for an n -type substrate with $N_D = 10^{17} \text{ cm}^{-3}$, an SiO_2 insulator with $t_{ox} = 60 \text{ \AA}$, and a polysilicon gate doped n -type with $N_D = 10^{20} \text{ cm}^{-3}$. Assume $Q'_o = 5 \times 10^{-9} \text{ C/cm}^2$.
- 2.3. Prove (2.5.14) and (2.5.15).
- 2.4. In Sec. 2.5.2 we have shown that, if ψ_s is assumed to be "pinned" to a constant, the plot of Q'_I vs. V_{GB} will be a straight line. Show that a pinned ψ_s is not a necessary condition for a straight-line plot (although it is sufficient); find the correct necessary condition.

- 2.5. Find an approximation for $\Delta\phi$ in (2.5.23) as follows. Show that (2.5.15) can be written as $\psi_s = 2\phi_F + \phi_t \ln \{[(V_{GB} - V_{FB} - \psi_s)^2/\gamma^2 - \psi_s]/\phi_t\}$. In the argument of the logarithm, replace ψ_s by the approximation $\psi_s = 2\phi_F$ from (2.5.22). This gives an expression in the form of (2.5.23), with $\Delta\phi$ weakly dependent on V_{GB} . Show that, for practical cases, where γ is between 0.2 and $2 \text{ V}^{1/2}$, and $V_{GB} - V_{FB} - 2\phi_F$ is up to several V, $\Delta\phi$ is several ϕ_t ; an average value is $6\phi_t$. Thus, a better estimate can be obtained if, in the above equation, ψ_s in the argument of the logarithm is replaced by $2\phi_F + 6\phi_t$.
- 2.6. For a device with $N_A = 2.5 \times 10^{16} \text{ cm}^{-3}$, $t_{ox} = 100 \text{ \AA}$, and $V_{FB} = -0.97 \text{ V}$ plot $\ln |Q'_I|$ vs. V_{GB} in weak inversion, using (a) (2.5.7) with (2.5.15); (b) (2.5.32) with (2.5.34) and (2.5.35); and (c) (2.5.42) with (2.5.43). Comment on the accuracy of the last two approaches.
- 2.7. For $N_A = 10^{17} \text{ cm}^{-3}$, $t_{ox} = 60 \text{ \AA}$, and $V_{FB} = -1 \text{ V}$, plot $Q'_I(V_{GB})$ using: (a) (2.5.7) with (2.5.15); (b) (2.5.26), from $V_{GB} = V_{M0} + 0.6 \text{ V}$ to $V_{GB} = 3 \text{ V}$. Comment on the accuracy of (2.5.26). Assume initially $\phi_0 = 2\phi_F + 6\phi_t$ and modify this value if necessary to decrease the error.
- 2.8. Rewrite the equations of this chapter for the case of n -type substrates.
- 2.9. Show that in depletion (with ψ_s larger than a few ϕ_t) and in weak inversion C'_{gb} is approximately given by

$$C'_{gb} = \frac{C'_{ox}}{\sqrt{1 + \frac{4}{\gamma^2} (V_{GB} - V_{FB})}}$$

- 2.10. Determine the value of the surface potential and the gate-substrate voltage at the onset of moderate inversion (ϕ_{M0} , V_{M0}), as well as the extrapolated threshold (V_{T0}) for the process of Example 2.2. Also, give rough estimates for the onset of strong inversion (ϕ_{H0} , V_{H0}).
- 2.11. Use (B.5) in Appendix B to prove (2.4.18). Note that the result is independent of the thickness of the shaded area containing Q_C in Fig. 2.4.
- 2.12. Equation (2.5.1) was derived as a special case of the general $Q'_C(\psi_s)$ relation in (2.4.17). Show that (2.5.1) can also be proved directly, using the several hints that follow. For $\psi(y)$ larger than a few ϕ_t , the hole charge is negligible, and $\rho(y) = -q[N_A + n(y)]$, with $n(y)$ related to $\psi(y)$ by $n(y) = N_A e^{(\psi - 2\phi_F)/\phi_t}$ in analogy with (2.4.12). Thus Poisson's equation (1.2.13) will be

$$\frac{d^2\psi}{dy^2} = \frac{qN_A}{\epsilon_s} \left(1 + e^{(\psi - 2\phi_F)/\phi_t}\right)$$

Multiply the above equation by $2(d\psi/dy)$, and recognize the resulting left-hand side as $(d/dy)(d\psi/dy)^2$. Integrate both sides from a point y_{bulk} outside the depletion region to a point y_{surface} at the surface; note that in the bulk, $\psi = 0$ and $d\psi/dy = 0$. Solve for $d\psi/dy$ at the surface, and then relate that quantity to Q'_C (see Appendix B). Solve for Q'_C and show that the result is practically the same as (2.5.1) for ψ_s larger than a few ϕ_t .¹³

- 2.13. Show that, independently of the charge in the inversion layer, if the thickness of this layer is assumed to be zero, the potential across the layer will also be zero (use the material in Appendix B).

- 2.14.** Use (B.6) in Appendix B to prove (2.5.5), assuming the “depletion approximation” (Sec. 1.5).
- 2.15.** Plots of charge density, electric field, and potential vs. distance from the two-terminal MOS structure are given in Appendix E. Verify these plots by using the basic laws of electrostatics of Appendix B and give values for critical points on the plots.
- 2.16.** (a) Show that, if a fixed charge sheet of charge per unit area Q' is located in the oxide at a distance d from the gate, its contribution to the flat-band voltage will be $-(1/C'_{\text{ox}})[(d/t_{\text{ox}})Q']$ (use material from Appendix B).
- (b) Show that, if a charge is distributed within the oxide uniformly along the horizontal dimension in Fig. 2.2, as described by a charge density per unit volume $\rho(y)$, its contribution to the flat-band voltage will be $-(1/C'_{\text{ox}})\int_{y_g}^{y_s}(y/t_{\text{ox}})q\rho(y) dy$, where y_g is y at the gate-oxide interface and y_s is y at the oxide-substrate interface.
- 2.17.** Plot the logarithm of the magnitude of the inversion layer charge, $\ln |Q'_I|$, vs. the surface potential ψ_s and verify that this plot approaches straight-line behavior in the weak and strong inversion regions. Use the process parameters of Prob. 2.6. Determine and compare the slopes of the plot in these two regions. (*Hint:* Show first that $d \ln |Q'_I|/d\psi_s = C'_i/|Q'_I|$).

CHAPTER 3

THE THREE- TERMINAL MOS STRUCTURE

3.1 INTRODUCTION

A complete MOS transistor is formed by adding two more terminals to the basic MOS structure of Chap. 2 to contact two opposite ends of the inversion layer. Through these terminals a potential difference can be applied across the inversion layer and a current can be caused to flow in it. A number of phenomena can then be observed, some directly associated with the current flow, some not. It is not convenient to introduce all these phenomena when they are present simultaneously. Some of these, specifically the ones not directly associated with current flow, can best be isolated and studied by themselves by means of a structure simpler than the MOS transistor, which we call a *three-terminal MOS structure*. This structure is formed by contacting the inversion layer of the basic MOS structure at only one end. We will study the changes that take place in the charges and the potential distribution of the three-terminal structure, caused by the application of an external voltage between this new terminal and the substrate. By the end of this short chapter, we will have all the facts necessary for a careful and convenient look at the MOS transistor, which will be the subject of all succeeding chapters.

3.2 CONTACTING THE INVERSION LAYER

Assume that an n^+ region is added to the basic MOS two-terminal structure, so that the structure shown in Fig. 3.1a is obtained; a constant V_{GB} will be assumed until further notice. Consider the n^+p junction formed by this region and the substrate. The depletion region on the p side contains ionized acceptor atoms as shown. The narrow depletion region part in the n^+ material containing ionized donor atoms (Fig. 1.15) is not shown for simplicity. As indicated in Sec. 1.5, one can short-circuit the n^+ region terminal to the substrate terminal without altering the picture, as shown in Fig. 3.1b. For this connection the part of the structure to the right of the n^+ region is still governed by the basic equations we have developed for the two-terminal structure, except for points close to the n^+ region. Such points can be affected directly by the two-dimensional field distribution around this region. However, we will assume for the present that the gate is long and wide, so that edge effects can be neglected over practically all the gate's length and width.

Let us now assume that V_{GB} is fixed at some value, producing a surface potential ψ_s underneath the oxide. From (2.5.7) and the associated discussion of the various regions of operation, we recall that ψ_s is "fighting against" $2\phi_F$. For example, if ψ_s is less than about $2\phi_F$, the exponential is very small in that equation and we are in weak inversion or in depletion. On the other hand, if ψ_s is greater than $2\phi_F$, the exponential term becomes important and we are in moderate or strong inversion.

We will now place a voltage source of value V_{CB} between the n^+ region and the substrate terminal, as shown in Fig. 3.1c. The value of V_{CB} will be assumed *nonnegative* to ensure that the n^+p junction is *not* forward-biased. This will correspond to the practical cases of interest when the three-terminal structure will become part of an MOS transistor in Chap. 4. If V_{CB} is not zero, communication of the inversion layer with the external voltage source (through the n^+ region) can drastically change the situation in comparison with the equilibrium case discussed in Chap. 2. The structure is now in *nonequilibrium*, and extra care should be exercised when attempting to apply relations from that chapter. As we will see, most such relations will have to be modified before they can apply to the present case.

Assume initially $V_{CB} = 0$, which makes the situation identical to that in Fig. 3.1b. Let us fix V_{GB} so that the surface potential is fixed at some value ψ_1 , bringing the surface to some level of inversion.[†] Consider now an increase in V_{CB} . The n^+ region will become more positive than before by V_{CB} . The positive potential will attract electrons from the inversion layer, which will flow toward the n^+ region, and from it into the top terminal of the voltage source. The inversion level will decrease; in fact, if V_{CB} is large enough, the inversion layer can disappear altogether. To restore the surface to its previous condition, the surface potential must be increased by the

[†]Note that, since the potential of the surface is ψ_1 with respect to the bulk, whereas that of the n^+ region is ϕ_{bi} (the built-in potential of the junction), there will be a region along the surface around the n^+p junction boundary where the potential changes from ϕ_{bi} to ψ_1 . Our results in this section will be valid to the right of this region.

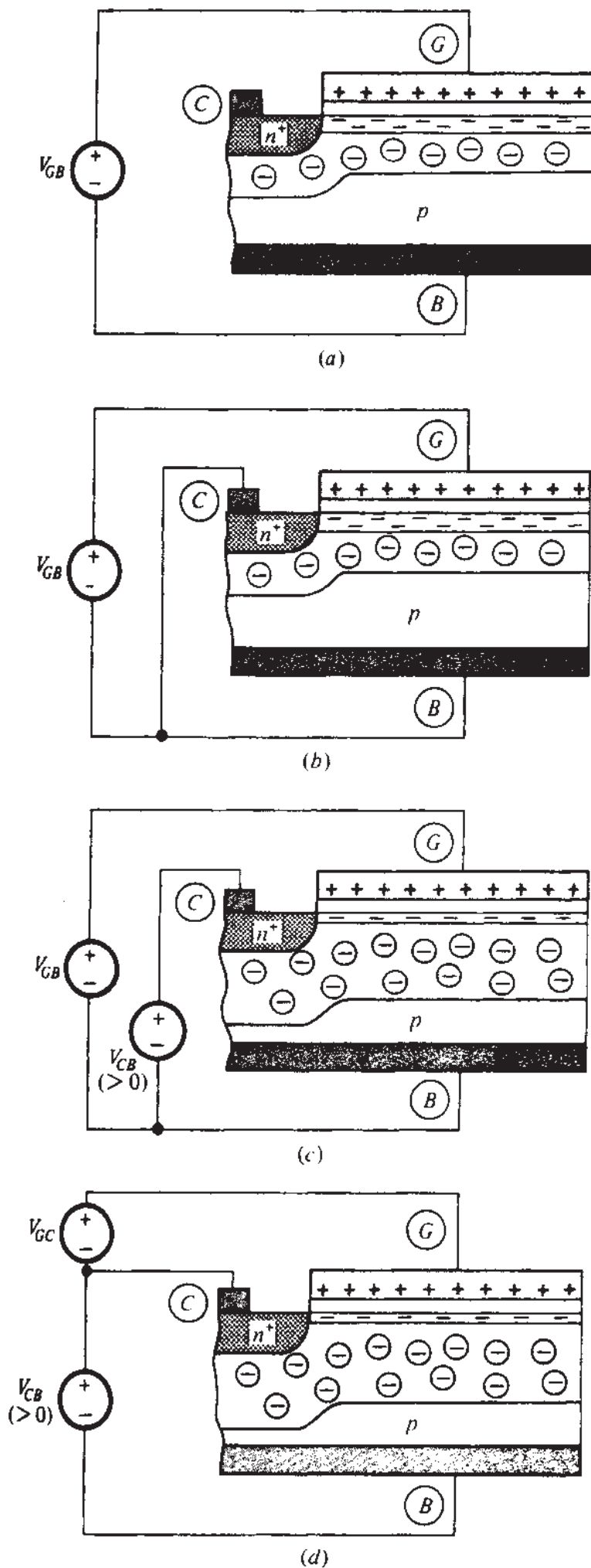


FIGURE 3.1

Three-terminal MOS structure, with n^+ region (a) left open; (b) short-circuited to p -type substrate; (c) biased at $V_{CB} (> 0)$ with respect to the substrate; (d) connection equivalent to (c) with voltages referred to terminal C.

same amount the potential of the n^+ region has increased, so that the latter is no longer more attractive for the electrons. Thus, the surface potential must be increased from ψ_1 to $\psi_1 + V_{CB}$. Then the surface will be at the original level of inversion again. To achieve such an increase in the surface potential, V_{GB} must be increased by an appropriate amount. We thus see that what determines the “attractiveness” of the surface for the electrons is not how large ψ_s is, but rather how large ψ_s is in comparison to V_{CB} . What counts is the *difference* $\psi_s - V_{CB}$. As long as that difference is fixed, the electron concentration at the surface is also fixed. The role of ψ_s in (2.4.12) is then played here by $\psi_s - V_{CB}$, and we have^{1,2†}

$$n_{\text{surface}} = N_A e^{[\psi_s - (2\phi_F + V_{CB})]/\phi_t} \quad (3.2.1)$$

Thus, just as in (2.4.12) ψ_s is “fighting against” $2\phi_F$ to increase the level of inversion at the surface, here ψ_s is instead fighting against $2\phi_F + V_{CB}$.

It should be noted that, with $V_{CB} > 0$, a small current will flow through the battery. This current is due partly to the junction reverse-bias current flowing from the n^+ region to the substrate (Sec. 1.5) and partly to a similar current flowing from the inversion layer to the substrate. The magnitude of the current is very small except at very high temperatures, and we will neglect both its electron and hole components in much of this book. It is important, though, to acknowledge its presence, as this current is a manifestation of the communication between the inversion layer and the external battery, and the resulting nonequilibrium condition.

The arguments leading to (3.2.1) cannot be extended to holes, as the n^+ region is not attractive to them. Thus the hole concentration will still be given by the same relations used for the two-terminal structure in Sec. 2.4. A general analysis including the effect of holes and electrons, and valid in accumulation, depletion, and inversion, is possible. It proceeds along the lines of the analysis in Sec. 2.4.4, only including the effect of V_{CB} on the mobile electron concentration n . Readers interested in such an analysis are referred to Appendix I. In the rest of this chapter we will concentrate on *inversion*, which will be defined by $n_{\text{surface}} \geq n_i$, where n_{surface} is the electron concentration at the surface and n_i is the intrinsic carrier concentration.[‡] In inversion, relatively simple results become possible. The considerations here parallel those for the two-terminal structure, so our discussion will be kept relatively brief.

The MOS system in inversion will still be characterized by five equations [as discussed for the two-terminal structure starting with (2.5.9)]. Of the five equations,

†In energy band parlance, V_{CB} serves to “split” the so-called “quasi-Fermi levels” of electrons and holes (Appendix H).

‡At the limit between depletion and inversion, $n_{\text{surface}} = n_i$. However, in contrast to the case of the two-terminal MOS structure, at this point the *hole* surface concentration is *not* equal to n_i , but is smaller than this quantity (assuming $V_{CB} > 0$). The interested reader can deduce this from the relations given in Appendix H or I.

four are identical to those for the two-terminal structure, and are repeated here for convenience:

$$V_{GB} = \psi_{ox} + \psi_s + \phi_{MS} \quad (3.2.2)$$

$$Q'_G + Q'_o + Q'_I + Q'_B = 0 \quad (3.2.3)$$

$$Q'_G = C'_{ox}\psi_{ox} \quad (3.2.4)$$

$$Q'_B = -\sqrt{2q\epsilon_s N_A} \sqrt{\psi_s} \quad (3.2.5a)$$

$$= -\gamma C'_{ox} \sqrt{\psi_s} \quad (3.2.5b)$$

where $\sqrt{2q\epsilon_s} = 5.8 \times 10^{-16} \text{ C} \cdot \text{cm}^{-1/2} \cdot \text{V}^{-1/2}$ and γ has been defined in (2.5.16). As in Chap. 2, the charge sheet³⁻⁸ and depletion approximations are implied in writing the above equation for Q'_B [see explanation above (2.5.5)].

To arrive at a complete set of equations, from which the behavior of the system can be determined, one more equation is needed, giving Q'_I as a function of ψ_s . However, (2.5.7) will not be valid as is, since it does not take into account the presence of V_{CB} . As might be expected from our discussion above, the correct relation for $Q'_I(\psi_s)$ is obtained simply by replacing $2\phi_F$ by $2\phi_F + V_{CB}$ in that equation (Appendix I or Prob. 3.12)³⁻⁸:

$$Q'_I = -\sqrt{2q\epsilon_s N_A} \left(\sqrt{\psi_s + \phi_i e^{[\psi_s - (2\phi_F + V_{CB})]/\phi_i}} - \sqrt{\psi_s} \right) \quad (3.2.6)$$

Other expressions for Q'_I have been discussed in the literature.⁹⁻¹⁴ From the above five equations everything else can be developed in the same way as for the two-terminal structure; thus, from (3.2.2) to (3.2.4) we have,

$$V_{GB} = V_{FB} + \psi_s - \frac{Q'_B(\psi_s) + Q'_I(\psi_s)}{C'_{ox}} \quad (3.2.7a)$$

where V_{FB} is the flat-band voltage defined in (2.2.6). Using (3.2.5b) and (3.2.6) in (3.2.7a), V_{GB} can be written as follows:

$$V_{GB} = V_{FB} + \psi_s + \gamma \sqrt{\psi_s + \phi_i e^{[\psi_s - (2\phi_F + V_{CB})]/\phi_i}} \quad (3.2.7b)$$

This equation cannot be solved explicitly for ψ_s . If V_{GB} and V_{CB} are given and ψ_s is desired, the equation can be solved numerically. The solution of this equation will be revisited in the context of the complete MOS transistor in Chap. 4.

The equation for Q'_I uses ψ_s and V_{CB} as independent variables. For future use it will be useful to express Q'_I also as a function of V_{GB} and ψ_s . This can be most easily done by using (3.2.2) to (3.2.5):

$$Q'_I = -C'_{ox} \left(V_{GB} - V_{FB} - \psi_s + \frac{Q'_B}{C'_{ox}} \right) \quad (3.2.8a)$$

or

$$Q'_I = -C'_{ox} \left(V_{GB} - V_{FB} - \psi_s - \gamma \sqrt{\psi_s} \right) \quad (3.2.8b)$$

Similarly, for the gate charge per unit area we obtain, from (3.2.2) and (3.2.4),

$$Q'_G = C'_{ox}(V_{GB} - V_{FB} - \psi_s) - Q'_o \quad (3.2.9)$$

Refer now to Fig. 3.1c, and assume that V_{CB} is held constant. We can then define a per-unit-area small-signal capacitance of the gate to the rest of the structure, $C'_g = dQ'_G/dV_{GB}$, and capacitances C'_c , C'_b , and C'_i as for the two-terminal structure. Reasoning as in Sec. 2.6, we obtain:

$$\frac{1}{C'_g} = \frac{1}{C'_{ox}} + \frac{1}{C'_c} \quad (3.2.10a)$$

or

$$\frac{1}{C'_g} = \frac{1}{C'_{ox}} + \frac{1}{C'_b + C'_i} \quad (3.2.10b)$$

where the expressions for C'_c , C'_b , and C'_i look like those in Sec. 2.6, except that $2\phi_F$ is replaced by $2\phi_F + V_{CB}$ (Appendix I):

$$C'_c = \sqrt{2q\epsilon_s N_A} \frac{1 + e^{[\psi_s - (2\phi_F + V_{CB})]/\phi_t}}{2\sqrt{\psi_s + \phi_t} e^{[\psi_s - (2\phi_F + V_{CB})]/\phi_t}} \quad (3.2.11)$$

$$C'_b = \sqrt{2q\epsilon_s N_A} \frac{1}{2\sqrt{\psi_s + \phi_t} e^{[\psi_s - (2\phi_F + V_{CB})]/\phi_t}} \quad (3.2.12)$$

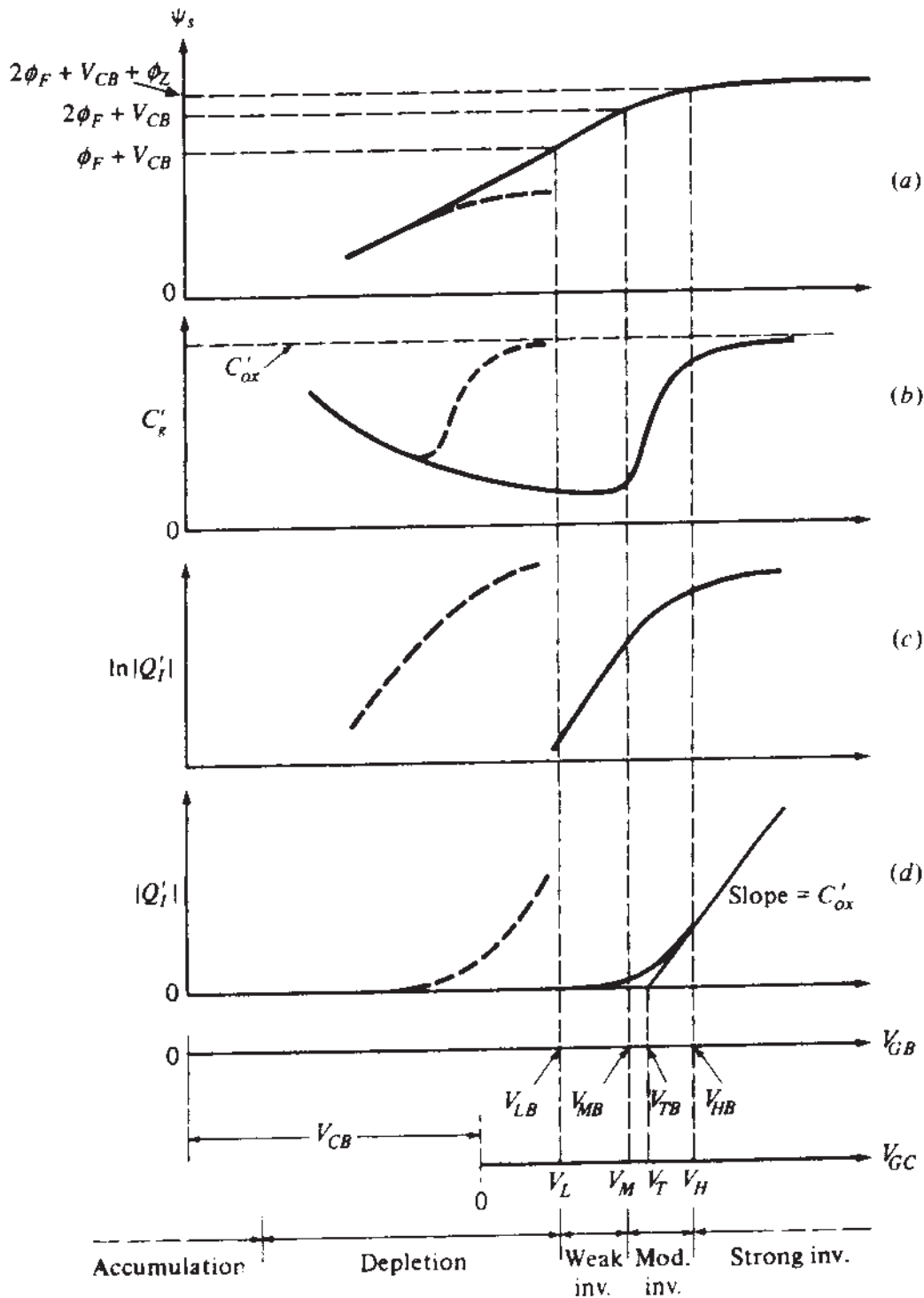
$$C'_i = \sqrt{2q\epsilon_s N_A} \frac{e^{[\psi_s - (2\phi_F + V_{CB})]/\phi_t}}{2\sqrt{\psi_s + \phi_t} e^{[\psi_s - (2\phi_F + V_{CB})]/\phi_t}} \quad (3.2.13)$$

As in Chap. 2, the expressions we give for C'_b and C'_i are more accurate than those that would be obtained by differentiating our expressions for Q'_B and Q'_I . We do this because (3.2.5) and (3.2.6), although adequate for calculating Q'_B and Q'_I , are not accurate enough to provide correct derivatives with respect to ψ_s [see discussion preceding (2.6.14) for details]. It is easy to note from (3.2.12) and (3.2.13) that $C'_i = C'_b$ at the point $\psi_s = 2\phi_F + V_{CB}$.

Plots of ψ_s , C'_g , $\ln|Q'_I|$, and $|Q'_I|$ vs. V_{GB} are shown in Fig. 3.2; the V_{GC} axis at the bottom will be considered later. The plots shown by the broken lines are for the case of Fig. 3.1b. These plots are the same as for a corresponding two-terminal structure, with one important exception for the C'_g plot. Let us compare this plot to Fig. 2.18. In the latter, the broken line represents C'_{gb} for the two-terminal structure at high frequencies of operation. This behavior was claimed to be different from that observed at low frequencies (solid line). The reason was traced to the fact that in the two-terminal structure the inversion layer is practically "isolated from the outside world." However, in the structure of Fig. 3.1b or Fig. 3.1c communication with the outside world is possible through the n^+ region. Thus, in strong inversion plenty of electrons are available just below the oxide, and their total charge "tracks" variations in V_{GB} even if these variations are of rather high frequency. Low-frequency and high-frequency behaviors then coincide, and only one curve is observed, as shown by the broken line in Fig. 3.2b. However, if an *extremely* high frequency is used, the supply of charge to the inversion layer might once again be unable to keep up; a behavior similar to that shown by the broken line in Fig. 2.18 could then be observed.

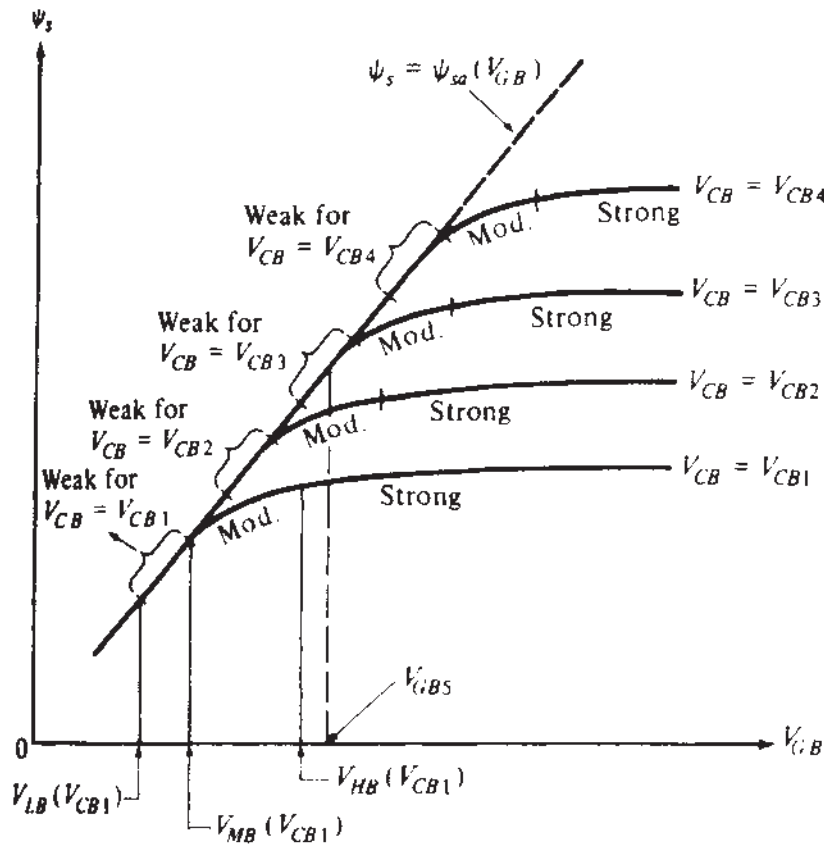
The plots given by the solid lines in Fig. 3.2 are for a given $V_{CB} > 0$. The behavior seen is qualitatively similar to that for $V_{CB} = 0$. Since the latter case is related to the two-terminal structure of Chap. 2, we can carry over several of the concepts and definitions discussed there to the general case of $V_{CB} > 0$. Thus, weak-, moderate-, and strong-inversion regions are again defined, as indicated at the bottom of the figure. The lower limits of these regions in terms of V_{GB} are denoted correspondingly by V_{LB} , V_{MB} , and V_{HB} as shown. Definitions for each region and values for the above quantities will be given later on. For now, it suffices to say that the regions are defined in such a way that in weak inversion the plot of $|Q'_I|$ vs. V_{GB} is essentially exponential; in strong inversion, it is essentially a straight line; and in moderate inversion, it is neither.

Note that the plots in Fig. 3.2 are for a given V_{CB} value. If V_{CB} is increased above that value, the plots will be qualitatively similar but the points where the effects of inversion become apparent will be shifted to the right, as predicted by the corresponding equations. To gain some further intuition about this effect, let us assume that in Fig. 3.1c V_{CB} is fixed at some value. Assume that V_{GB} is so low that not only the inversion layer is absent but also the depletion region width is smaller than shown. Let us now increase V_{GB} . Doing so will not change V_{CB} , the reverse bias of the n^+p junction on the left. Hence the depletion region under the n^+ region will not be affected. However, more positive charge will now be placed on the gate, which must be balanced by more negative charge under the oxide. An increase in the depletion region width accomplishes this balance. If V_{GB} is increased far enough, a point will eventually be reached where the depletion region under the oxide will be nearly as wide as that under the n^+

**FIGURE 3.2**

Various quantities characterizing the structure of Fig. 3.1c plotted vs. V_{GB} and V_{GC} . (a) Surface potential; (b) gate capacitance to the rest of the structure per unit area; (c) logarithm of inversion layer charge magnitude per unit area; (d) inversion layer charge magnitude per unit area. For the broken lines $V_{CB} = 0$; the solid lines are for a V_{CB} of a given positive value.

region; the value of V_{GB} needed for this to happen depends on the value of V_{CB} . When this happens, the potential across the depletion regions under the oxide and under the n^+ region will be about the same. Thus, a point at the surface is at about the same potential (with respect to the substrate deep in the bulk) as is a point in the n^+ region. Now the surface is about as "attractive" for electrons as is the n^+ region. Electrons are

**FIGURE 3.3**

Surface potential vs. gate-substrate voltage with V_{CB} as a parameter. For each V_{CB} the regions of inversion are indicated on the corresponding curve.

attracted most to regions with the most positive potential, and they have now no reason to prefer only the n^+ region as opposed to the “surface” on its right.[†] An inversion layer is thus formed. If now V_{CB} is raised further, the depletion region under the n^+ region will become again wider than under the inversion layer. The electrons will tend again to favor the n^+ region rather than the surface, and the inversion layer will tend to disappear. If it is desired to restore the level of inversion previously achieved, V_{GB} will have to be raised further to make the depletion region width under the gate approximately the same as that under the n^+ region once again. Then the potential at the surface will again be about the same as the potential of the n^+ region (both with respect to the bulk), and electrons can once more be attracted to the surface. From the above picture, the competing roles of V_{CB} and V_{GB} are clear. Increasing V_{CB} tends to make the level of inversion lighter; increasing V_{GB} tends to make that level heavier.

It is interesting to investigate the effect of these phenomena on the surface potential. In Fig. 3.3 we show ψ_s vs. V_{GB} , as it results from (3.2.7b), for different values

[†]To be more quantitative than this, one must recall that the potential at the surface is ψ_s , whereas that of the n^+ region is $\phi_{bi} + V_{CB}$, with ϕ_{bi} the built-in potential of the n^+p junction (Sec. 1.5). It can be verified by using (3.2.1) that, to make the electron concentration at the surface equal to that in the n^+ region, requires $\psi_s = \phi_{bi} + V_{CB}$. For other values of ψ_s , the surface concentration will be different, as determined by (3.2.1), and there will be a transition region along the surface around the n^+p junction boundary, over which the potential changes from $\phi_{bi} + V_{CB}$ to ψ_s . As already mentioned, we are focusing on the part of the structure to the right of this transition region. As one goes from weak toward strong inversion, the potential across this horizontal transition region diminishes.

of V_{CB} . The regions of inversion indicated can be interpreted in terms of the corresponding Q'_I plots (which are not shown, but are nevertheless nearly exponential in weak, nearly straight lines in strong, and neither in moderate inversion). Note that increasing V_{CB} "postpones" the tendency of ψ_s to "flatten out" until larger V_{GB} values. This is because it "postpones" the formation of an inversion layer, as already explained. Assume now that, for a given V_{CB} , the value of V_{GB} is so low that the inversion layer charge is negligible (e.g., in depletion); denote the corresponding value of ψ_s by ψ_{sa} . Then (3.2.7a) gives

$$V_{GB} \approx V_{FB} + \psi_{sa} - \frac{Q'_B(\psi_{sa})}{C'_{ox}} \quad (3.2.14)$$

which, using (3.2.5b), becomes:

$$V_{GB} \approx V_{FB} + \psi_{sa} + \gamma \sqrt{\psi_{sa}} \quad (3.2.15)$$

Solving this, we obtain ψ_{sa} as a function of V_{GB} :

$$\psi_{sa}(V_{GB}) \approx \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right)^2 \quad (3.2.16)$$

This function has already been encountered in Sec. 2.5.3, and is shown by the broken curve in Fig. 3.3. It can be seen that for any V_{CB} , as long as $\psi_s \approx \psi_{sa}(V_{GB})$, the effect of V_{CB} on ψ_s is negligible. Physically, this is because in such a case there is practically no inversion layer charge. Most electric field lines coming from the gate "pass right through" the practically absent inversion layer and terminate on ionized acceptor atoms on the substrate. Thus Q'_I has "no handle" on the overall balance of the system represented by (3.2.3), and, although varying V_{CB} will vary drastically the negligible $|Q'_I|$ in a relative sense, the rest of the system simply "does not notice." Note that $|Q'_I|$ will be negligible as long as the exponential in (3.2.6) is negligible. In fact, the same exponential appears in the general $V_{GB}(\psi_s)$ relation (3.2.7b). When the exponential there is negligible, this equation reduces to (3.2.15). Things will begin thus changing significantly when the exponential starts becoming important, which will happen at about $\psi_s \approx 2\phi_F + V_{CB}$. Then the complete (3.2.7b) must be used to find ψ_s , and the plot for the latter will begin to flatten out, as seen in Fig. 3.3.

It can be seen from the above discussion and Fig. 3.3 that $\psi_{sa}(V_{GB})$ is the surface potential that develops, for a given V_{GB} , if the inversion layer is prevented from forming (by maintaining V_{CB} sufficiently high). As seen in Fig. 3.3, the slope of $\psi_{sa}(V_{GB})$ is a weak function of V_{GB} ; this fact will be seen to lead to certain simplifications later on. Consistent with Sec. 2.5.3, we will denote the inverse of this slope by n :

$$n \equiv \left(\frac{d\psi_{sa}}{dV_{GB}} \right)^{-1} \quad (3.2.17)$$

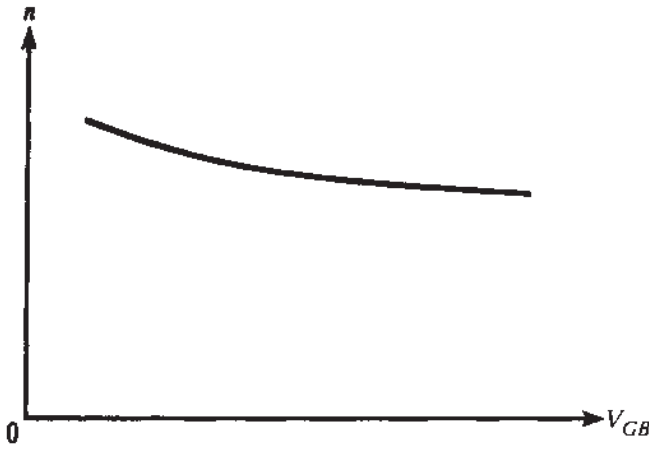


FIGURE 3.4

The parameter n as a function of the gate-substrate voltage.

Using (3.2.16) in the above definition, we obtain, as in Sec. 2.5.3:

$$n = 1 + \frac{\gamma}{2\sqrt{\psi_{sa}(V_{GB})}} \quad (3.2.18)$$

Typical values for n are between 1 and 1.5. A plot of n vs. V_{GB} is shown in Fig. 3.4.

Now let V_{GB} have a given value, say V_{GB5} as shown in Fig. 3.3 (the subscript 5 is used in connection with a discussion later in this chapter). It is clear that one cannot state the level of inversion, knowing only this value of V_{GB} ; rather, the structure can be in strong, moderate, or weak inversion, depending on the value of V_{CB} , as seen in the figure.

Before closing this section, we should note that in some of our discussions we will find it convenient to use the connection shown in Fig. 3.1d rather than the one in Fig. 3.1c. Note that since

$$V_{GB} = V_{GC} + V_{CB} \quad (3.2.19)$$

the plots of Fig. 3.2 can be viewed as being versus V_{GC} simply by shifting the 0 point of the horizontal axis by V_{CB} . This is shown by using a second horizontal axis at the bottom of the figure. The symbols V_L , V_M , and V_H denote the onset of weak, moderate, and strong inversion, respectively, in terms of V_{GC} .

3.3 THE BODY EFFECT

From the qualitative discussion in the previous section, it is clear that increasing V_{CB} decreases the level of inversion, unless we also increase V_{GB} by an appropriate amount. It turns out that, if we want to keep the level of inversion the same, the increase in V_{GB} must be *larger* than the increase in V_{CB} . This is referred to as the *body effect* or *substrate effect*. In this section, we will attempt to provide intuition about it. Quantitative results will follow in subsequent sections, where we consider the limits of regions of inversion.

It is best to discuss the effect mentioned with the help of Fig. 3.1*d*. Body effect really refers to the fact that, if V_{CB} in the connection shown is raised while V_{GC} is kept constant, then the level of inversion will decrease, although $V_{GB} = V_{GC} + V_{CB}$ is increased, and that, if we want to restore the original level of inversion, we will have to increase V_{GC} . To see the reason why, let us assume first that V_{GC} is large enough to cause *strong* inversion. The strongly inverted surface, containing an abundance of electrons, is sometimes likened to a n^+p region. In fact, the inversion-layer-substrate combination of this case is sometimes called a *field-induced n^+p junction*. In several respects this junction has a similar behavior as the regular n^+p junction discussed in Sec. 1.5. The role of the reverse bias V_R there is played here by V_{CB} . Thus, V_{CB} will be called the *effective reverse bias* of the field-induced junction in the strong-inversion region. Just like increasing V_R widens the depletion region in a regular junction, increasing V_{CB} will widen the depletion region under the strongly inverted surface. Assume now that we are well into strong inversion, and that V_{CB} in Fig. 3.1*d* is increased by a small amount. In the connection shown in this figure, this will not affect V_{GC} . The strong-inversion layer, with its abundance of electrons, can be viewed as the bottom plate of a capacitor, with the gate being the top plate. Since V_{GC} has not changed, the voltage across this "capacitor" will remain essentially unchanged, and there will be practically no change in the gate charge.

Since V_{CB} has increased, the depletion region will widen, more acceptor atoms will be uncovered, and the total depletion region charge will become more negative. The *total* negative charge under the oxide, however, must balance the positive gate charge, which has remained practically unchanged, as already explained. Since more ionized acceptor atoms are now available in the depletion region to contribute to this "balancing" function, fewer free electrons are needed in the inversion layer. Thus, the level of inversion will decrease. If it is desired to restore the previous level of inversion, one must then increase V_{GC} by a sufficient amount.

The larger the substrate doping, the larger the change in the depletion region charge for a given change in V_{CB} [this can be easily deduced by analogy to (1.5.13) and (1.5.11)]. Continuing the above argument, this means that a larger increase in the gate voltage will be needed then to restore the original level of inversion. Thus, the body effect is more pronounced for heavily doped substrates. This effect will also be more pronounced for thicker oxides. The thicker the oxide, the weaker the influence of the gate on the inversion layer charge, and the larger the gate voltage needed to influence that charge to a given degree. The above dependence on substrate doping and oxide thickness is manifest in the expression for the *body effect coefficient* γ in (2.5.16), which will be seen to enter directly in quantitative results for the body effect in Sec. 3.4.1.

The above arguments (likening the inversion-layer-substrate structure to a n^+p junction, viewing V_{CB} as its "effective reverse bias," and considering the inversion layer as the bottom plate of a capacitor) are *only valid in strong inversion*. In moderate and, especially, in weak inversion, the electron concentration in the inversion layer is light, and the layer *cannot* be viewed as an n^+ region or as a capacitor plate. A significant part of the electric field lines from the gate pass through the inversion layer and terminate directly on ionized acceptor atoms in the depletion region. In

fact, as already explained in Sec. 3.2, in weak inversion, varying V_{CB} will leave the surface potential practically unaffected (see Fig. 3.3), and so cannot affect appreciably the depletion region width. Concerning, however, the effect of V_{CB} on the level of inversion, the same qualitative conclusions as for strong inversion can be reached. Thus, assume that we are in weak inversion and that V_{CB} in Fig. 3.1d is increased. To keep the level of inversion the same, ψ_s must increase by about the same amount, as has already been argued in Sec. 3.2. To increase ψ_s by this amount, V_{GB} must be increased by a larger amount, as seen from (3.2.7b); thus, $V_{GC} = V_{GB} - V_{CB}$ must be increased.

As a result of the above effects, an increase in V_{CB} will result in increases in the values of V_L , V_M , and V_H marked on the V_{GC} axis in Fig. 3.2. The quantity V_T shown there, which is simply the "extrapolated threshold voltage" (in analogy to V_{T0} for the two-terminal structure), is also found to increase (Sec. 3.4.2). In fact, the term *body effect* is commonly taken to mean just the increase of V_T with V_{CB} .

3.4 REGIONS OF INVERSION

3.4.1 Approximate Limits

For the two-terminal MOS structure we have defined the onset of weak, moderate, and strong inversion by the surface potential values ϕ_F , $2\phi_F$, and $2\phi_F + \phi_{Z0}$ (see Fig. 2.9). For the three-terminal structure, on the basis of the discussion in Sec. 3.2, it is reasonable to define these onsets as indicated in the top row of Table 3.1. In the right-most entry, ϕ_Z is the width of the moderate inversion region in terms of surface potential; its value is several ϕ_F . The onset of strong inversion is chosen so that in this region $Q'_i(V_{GB})$ is essentially a straight line (see below). With $V_{CB} = 0$, the surface potential bounds shown reduce to their corresponding values for the two-terminal MOS structure.

The onsets of weak inversion and of moderate inversion in terms of V_{GB} (see Fig. 3.3) can be found by using the corresponding surface potentials in (3.2.7b) and neglecting the exponential term. Thus, for example, we have

$$V_{MB} = V_{FB} + (2\phi_F + V_{CB}) + \gamma\sqrt{2\phi_F + V_{CB}} \quad (3.4.1)$$

The corresponding values in terms of V_{GC} can be found by subtracting V_{CB} (see the axes at the bottom of Fig. 3.2). Thus, for example:

$$V_M = V_{MB} - V_{CB} \quad (3.4.2a)$$

$$V_M = V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F + V_{CB}} \quad (3.4.2b)$$

These results are shown in the second and third rows of Table 3.1 along with expressions for the other limit points. An ambiguity is involved in defining the onset of strong inversion, for the same reasons as those discussed in Sec. 2.5.2. Thus, approximate results are shown in Table 3.1. For more accurate results,¹⁵ the reader is

TABLE 3.1
Approximate bounds between regions†

| | Bound between depletion and weak inversion | Bound between weak and moderate inversion | Bound between moderate and strong inversion |
|--|--|---|---|
| In terms of surface potential ψ_s | $\phi_F + V_{CB}$ | $2\phi_F + V_{CB}$ | $2\phi_F + V_{CB} + \phi_L^\ddagger$ |
| In terms of V_{CB} , for a given V_{CB} | $V_{LB} = V_L + V_{CB}$ | $V_{MB} = V_M + V_{CB}$ | $V_{HB} = V_H + V_{CB}$ |
| In terms of V_{CB} , for a given V_{CB} | $V_L =$ $V_{FB} + \phi_F + \gamma\sqrt{\phi_F + V_{CB}}$ | $V_M =$ $V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F + V_{CB}}$ | $V_H = V_M + V_Z^\S$ |
| In terms of V_{CB} , for a given V_{CB} (see Sec. 3.5) | $V_U =$ $\left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}}\right)^2 - \phi_F$ | $V_W =$ $\left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}}\right)^2 - 2\phi_F$ | $V_Q =$ $\left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB} - V_Z}\right)^2 - 2\phi_F$ |

†For more accurate results, see Appendix G.

‡ ϕ_L is a weak junction of process parameters, temperature, and V_{CB} . Its value is several ϕ_F . One should *not* attempt to use an approximate value from this expression for ψ_s in expressions like (3.2.7b), as a small error in ψ_s can cause a large error in the exponential term.

§ V_Z is typically 0.5 to 0.6 V at room temperature, for practical combinations of substrate doping and oxide thickness, and for values of V_{CB} up to a few volts. It decreases somewhat with increasing V_{CB} .

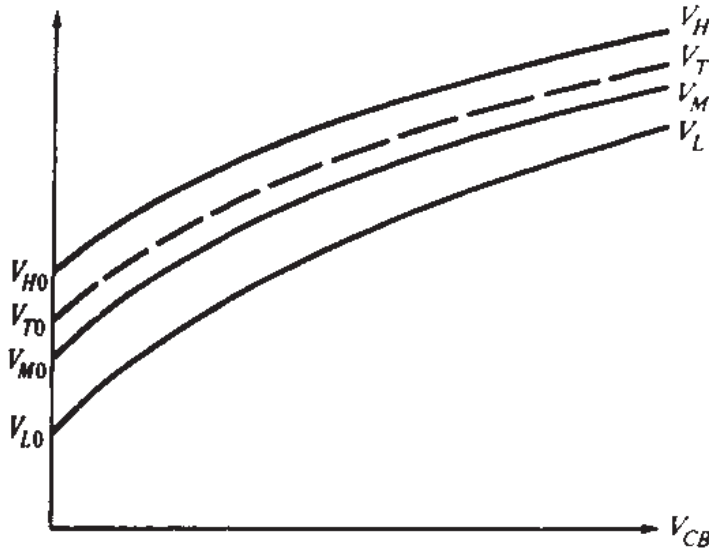


FIGURE 3.5

Onset of strong inversion V_H , extrapolated threshold V_T , onset of moderate inversion V_M , and onset of weak inversion V_L vs. V_{CB} for the three-terminal MOS structure of Fig. 3.1d.

referred to Appendix G. Discussion of the last row of Table 3.1 will be postponed until Sec. 3.5.

Plots of V_L , V_M , and V_H vs. V_{CB} for a given fabrication process look as shown in Fig. 3.5 (also shown is the plot for V_T , for which an expression will be derived in Sec. 3.4.2). The values of these quantities at $V_{CB} = 0$ are denoted by adding the subscript 0 as shown in the figure. These values are the same as for the two-terminal MOS structure, for reasons already discussed. The fact that the quantities in Fig. 3.5 increase with increasing V_{CB} is a consequence of the body effect discussed in Sec. 3.3. As is apparent from the expressions in Table 3.1, how much V_L , V_M , and V_H will increase for a given increase in V_{CB} is determined by the value of the coefficient γ ; hence, the name *body effect coefficient* for this quantity. We remind the reader that γ was defined in (2.5.16), which is repeated below:

$$\gamma = \frac{\sqrt{2q\epsilon_s N_A}}{C'_{ox}} \quad (3.4.3)$$

This quantity has been plotted in Fig. 2.8. It is seen that our intuitive predictions of Sec. 3.3 (that the body effect is stronger for heavier substrate dopings and/or thicker oxides) are verified by the entries for V_L , V_M , and V_H in Table 3.1, in conjunction with (3.4.3).

The definitions and properties of the regions of inversion in terms of ψ_s , V_{GB} , V_{GC} , and V_{CB} are summarized in Table 3.2. (The fourth row of the table will be discussed in Sec. 3.5.) The properties listed correspond to those for the two-terminal MOS structure, which were summarized in Table 2.1. Some of these properties will be discussed in the following sections, in which we look separately at each region of inversion. Since the considerations in each region are similar to those for the two-terminal structure of Chap. 2, we will be relatively brief to avoid boredom.

TABLE 3.2
Regions of inversion and properties (three-terminal MOS structure)

| | Weak inversion | Moderate inversion | Strong inversion |
|---|---|--|--|
| Definition in terms of surface potential ψ_s (see Fig. 3.2a) | $\phi_F + V_{CB} \leq \psi_s < 2\phi_F + V_{CB}$ | $2\phi_F + V_{CB} \leq \psi_s < 2\phi_F + V_{CB} + \phi_Z$ | $2\phi_F + V_{CB} + \phi_Z \leq \psi_s$ |
| Definition in terms of V_{GB} for a given V_{CB} (see Figs. 3.1c and 3.2) | $V_{LB} \leq V_{GB} < V_{MB}$ | $V_{MB} \leq V_{GB} < V_{HB}$ | $V_{HB} \leq V_{GB}$ |
| Definition in terms of V_{GC} for a given V_{CB} (see Figs. 3.1d and 3.2) | $V_L \leq V_{GC} < V_M$ | $V_M \leq V_{GC} < V_H$ | $V_H \leq V_{GC}$ |
| Definition in terms of V_{CB} for a given V_{GB} (see Figs. 3.1c and Sec. 3.5)† | $V_U \geq V_{CB} > V_W$ | $V_W \geq V_{CB} > V_Q$ | $V_Q \geq V_{CB}$ |
| $\frac{ Q'_I }{ Q'_B }$ | $\ll 1$ | Varies | $\gg 1$ deep in strong inversion; not necessarily so near the bottom of the region |
| $\frac{C'_I}{C'_b}$ | $\ll 1$ deep in weak inversion; not necessarily so near the top of the region | Varies | $\gg 1$ |
| $\frac{d\psi_s}{dV_{GB}}$ | Approximately constant; attains its maximum value in this region | Varies | Small |
| $\frac{d\psi_s}{dV_{CB}}$ | Very small | Varies | Close to 1 |
| Dependence of Q'_I on V_{GB} or V_{GC} for V_{CB} constant | Approximately exponential | — | Approximately first-degree polynomial |
| $\frac{d \ln Q'_I }{d\psi_s}$ | $\frac{1}{\phi_i}$ | Varies | $\frac{1}{2\phi_i}$ |

†It is assumed here that V_{GB} is sufficiently large that V_Q , V_W , and V_U come out non-negative; see Sec. 3.5.1 for the interpretation of negative values for these quantities.

3.4.2 Strong Inversion

As seen in Fig. 3.2, for a given V_{CB} strong inversion is defined by

$$V_{GB} \geq V_{HB}(V_{CB}) \quad (3.4.4)$$

Exact calculations show that, although deep in strong inversion $|Q'_I| \gg |Q'_B|$, for values of V_{GB} close to V_{HB} we can have Q'_I larger or smaller than Q'_B , depending

on substrate doping concentration and oxide thickness. What really makes the structure in strong inversion behave as it does is not the relative magnitude of $|Q'_i|$ and $|Q'_b|$ but rather that of their slopes with respect to ψ_s ; these slopes are simply C'_i and C'_b . One finds that, in strong inversion,

$$C'_i \gg C'_b \quad (3.4.5)$$

As seen in Fig. 3.2a, ψ_s changes only slightly with V_{GB} in strong inversion and can be assumed "pinned" to a fixed value. With $V_{CB} = 0$, that value is the same as that corresponding to the two-terminal structure, which was denoted by ϕ_0 in (2.5.21). Since the pinned value is close to $2\phi_F + V_{CB}$, it roughly follows the dependence of this quantity on V_{CB} . Thus the value of the pinned surface potential in strong inversion can be approximated by

$$\boxed{\psi_s \approx \phi_0 + V_{CB}} \quad (3.4.6)$$

where

$$\phi_0 \approx 2\phi_F + \Delta\phi \quad (3.4.7)$$

with $\Delta\phi$ equal to several ϕ_t (see the related discussion in Sec. 2.5.2).

The depletion region width can then also be assumed pinned at a value d_{Bm} which, corresponding to (2.5.24), is given by

$$d_{Bm} = \sqrt{\frac{2\epsilon_s}{qN_A}} \sqrt{\phi_0 + V_{CB}} \quad (3.4.8)$$

and the depletion region charge, from (3.2.5) and (3.4.6), is pinned at the value

$$Q'_B = -\sqrt{2q\epsilon_s N_A} \sqrt{\phi_0 + V_{CB}} \quad (3.4.9)$$

or, using (3.4.3):

$$\boxed{Q'_B = -\gamma C'_{ox} \sqrt{\phi_0 + V_{CB}}} \quad (3.4.10)$$

As already explained in Sec. 3.3, *for the case of strong inversion only*, V_{CB} can be interpreted as the effective reverse bias of the field-induced junction formed by the inversion layer and substrate. Changes in the potential between terminals C and B in Fig. 3.1c cause nearly equal changes in the potential ψ_s between the inversion layer and the bulk, as long as strong inversion is maintained. This is why in strong inversion $d\psi_s/dV_{CB} \approx 1$, as follows from (3.4.6).

The inversion layer charge can be obtained from (3.2.2) to (3.2.4):

$$Q'_i = -C'_{ox}(V_{GB} - \phi_{MS} - \phi_0 - V_{CB}) - Q'_o - Q'_B \quad (3.4.11a)$$

or

$$Q'_I = -C'_{ox}(V_{GB} - V_{TB}) \quad (3.4.11b)$$

where

$$V_{TB}(V_{CB}) = \phi_{MS} - \frac{Q'_o}{C'_{ox}} + \phi_0 + V_{CB} - \frac{Q'_B}{C'_{ox}} \quad (3.4.12a)$$

This, using (2.2.6), can be written as

$$V_{TB}(V_{CB}) = V_{FB} + \phi_0 + V_{CB} - \frac{Q'_B}{C'_{ox}} \quad (3.4.12b)$$

or, using (3.4.10):

$$V_{TB} = V_{FB} + \phi_0 + V_{CB} + \gamma\sqrt{\phi_0 + V_{CB}} \quad (3.4.12c)$$

The quantity $V_{TB}(V_{CB})$ is the *G-B (gate-substrate) extrapolated threshold voltage*; its meaning is illustrated in Fig. 3.2d. This quantity is related to the *G-C extrapolated voltage*[†] V_T (illustrated in the same figure) by:

$$V_{TB} = V_T + V_{CB} \quad (3.4.13)$$

Accordingly, we have

$$V_T = V_{FB} + \phi_0 - \frac{Q'_B}{C'_{ox}} \quad (3.4.14a)$$

or

$$V_T = V_{FB} + \phi_0 + \gamma\sqrt{V_{CB} + \phi_0} \quad (3.4.14b)$$

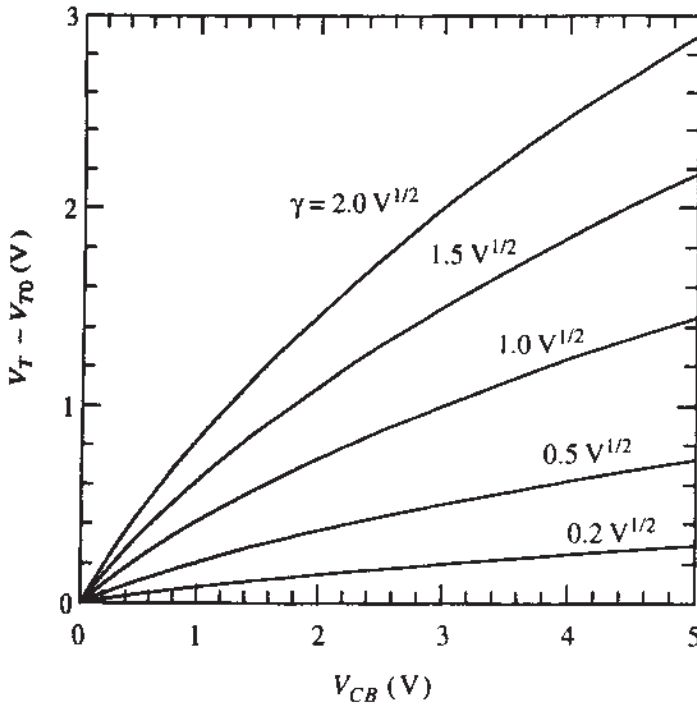
which can be written as

$$V_T = V_{T0} + \gamma(\sqrt{V_{CB} + \phi_0} - \sqrt{\phi_0}) \quad (3.4.15)$$

where V_{T0} is the value of V_T at $V_{CB} = 0$, given by

$$V_{T0} = V_{FB} + \phi_0 + \gamma\sqrt{\phi_0} \quad (3.4.16)$$

[†]C refers to terminal C in Fig. 3.1d.

**FIGURE 3.6**

Increase of the extrapolated threshold voltage V_T above its value at $V_{CB} = 0$, shown vs. V_{CB} for various values of the body effect coefficient γ . A single value of $\phi_0 = 1$ V is assumed.

which is, of course, the same quantity as that encountered for the two-terminal structure in (2.5.28).

In some treatments, as we have seen, $\phi_0 = 2\phi_F$ is assumed. In that case we have, by comparing (3.4.14b) to (3.4.2b):

$$V_T = V_M \quad \text{if } \phi_0 = 2\phi_F \text{ is used} \quad (3.4.17)$$

However, as has already been remarked (see Sec. 2.5.2), $2\phi_F$ is in general not the best choice for ϕ_0 .

The plot of V_T vs. V_{CB} has been included in Fig. 3.5. The threshold increase $V_T - V_{T0}$ due to the body effect is shown vs. V_{CB} for various values of γ in Fig. 3.6.†

Using $V_{GB} = V_{GC} + V_{CB}$ and $V_{TB} = V_T + V_{CB}$ in (3.4.11b), we obtain

$$Q'_I = -C'_{ox} (V_{GC} - V_T) \quad (3.4.18)$$

This corresponds to the straight-line part of the plot in Fig. 3.2d.

3.4.3 Weak Inversion

For a given V_{CB} , we define the weak-inversion region by (Fig. 3.2)

$$V_{LB}(V_{CB}) \leq V_{GB} < V_{MB}(V_{CB}) \quad (3.4.19)$$

†Plots of $V_L - V_{L0}$, $V_M - V_{M0}$, and $V_H - V_{H0}$ have similar shapes.

In this region, we have

$$Q'_I \ll Q'_B \quad (3.4.20)$$

Consider now (3.2.6). The magnitude of the second term under the first square root is small, because in weak inversion $\psi_s < 2\phi_F + V_{CB}$ (Fig. 3.2). Thus we can use an expansion like the one that led to (2.5.32) for the two-terminal structure to obtain

$$Q'_I = -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\psi_s}} \phi_t e^{[\psi_s - (2\phi_F + V_{CB})]/\phi_t} \quad (3.4.21)$$

We remind the reader that in weak inversion V_{CB} *cannot* be interpreted as an effective reverse bias, as explained in Sec. 3.3. In weak inversion, the surface potential is practically independent of V_{CB} and is practically equal to ψ_{sa} (see Fig. 3.3). The expression for ψ_{sa} was given in (3.2.16). Thus:

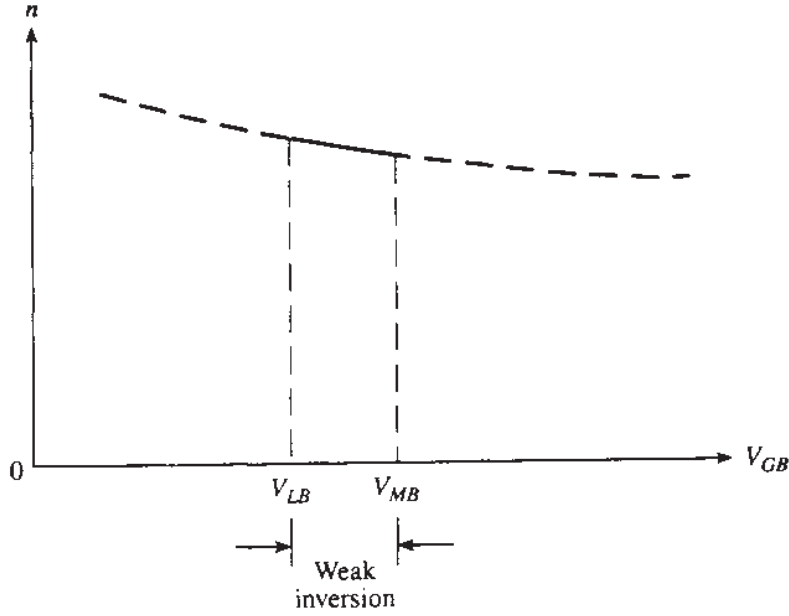
$$\psi_s \approx \psi_{sa}(V_{GB}) = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right)^2 \quad (3.4.22)$$

Therefore, in (3.4.21) the *only* term dependent on V_{CB} is $\exp(-V_{CB}/\phi_t)$. To emphasize this important conclusion, we write that equation as follows:

$$Q'_I = \underbrace{-\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\psi_{sa}(V_{GB})}} \phi_t e^{[\psi_{sa}(V_{GB}) - 2\phi_F]/\phi_t}}_{\text{dependent only on } V_{GB}} \cdot \underbrace{e^{-V_{CB}/\phi_t}}_{\text{dependent only on } V_{CB}} \quad (3.4.23)$$

For a fixed V_{CB} , Q'_I turns out to be nearly exponentially dependent on V_{GB} . To see this, we now develop an expression for Q'_I which does not contain ψ_{sa} , as we did for the two-terminal structure. In Fig. 3.7, we have repeated one of the curves of Fig. 3.3. Let V'_{CB} denote the constant value of V_{CB} for which this curve is obtained. In the figure we show the weak-inversion limits on the surface potential from Table 3.1. As seen, the width of the region in terms of ψ_{sa} is only ϕ_F . As V_{GB} changes over the region, the corresponding variation of the term $\sqrt{\psi_{sa}}$ in (3.4.23) is very small compared to the drastic variation of the exponential in that equation. Thus we can assume^{16,17} that $\sqrt{\psi_{sa}}$ is practically fixed at $\sqrt{2\phi_F + V'_{CB}}$ with $2\phi_F + V'_{CB}$ being the value of ψ_{sa} at the top of the weak-inversion region, as shown by point *M* in Fig. 3.7.† Thus (3.4.23) becomes

†A more appropriate point¹⁸ would have been $1.5\phi_F + V_{CB}$, but the final results would be more complicated. At the level of approximation we are seeking, the point $2\phi_F + V_{CB}$ is acceptable.

**FIGURE 3.8**

The parameter n as a function of the gate-substrate voltage; over the weak inversion region, this parameter does not vary much.

Taking changes below point M in Fig. 3.7, we can write:

$$\psi_{sa} - (2\phi_F + V'_{CB}) \approx \frac{1}{n} (V_{GB} - V_{MB}) \quad (3.4.26a)$$

$$= \frac{1}{n} (V_{GC} - V_M) \quad (3.4.26b)$$

where V_{MB} and V_M are given in Sec. 3.4.1, and n is evaluated at point M :

$$n = 1 + \frac{\gamma}{2\sqrt{2\phi_F + V'_{CB}}} \quad (3.4.27)$$

Substituting (3.4.26) in (3.4.24), we obtain

$$Q'_I \approx Q'_M e^{(V_{GB} - V_{MB})/(n\phi_t)} \quad (3.4.28a)$$

or

$$Q'_I \approx Q'_M e^{(V_{GC} - V_M)/(n\phi_t)} \quad (3.4.28b)$$

where

$$Q'_M = -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{2\phi_F + V'_{CB}}} \phi_t \quad (3.4.29)$$

represents the value of Q'_I at the top of weak inversion ($V_{GB} = V_{MB}$). Equation (3.4.28) predicts exponential behavior for Q'_I in weak inversion. As was the case with the two-terminal structure, this is only approximately true. In cases where the slope of Q'_I vs. V_{GB} is important, this approximation may not be adequate.

If V_{CB} is fixed at a value V'_{CB} and V_{GB} is varied, (3.4.28a) can be a very helpful equation since it makes explicit the exponential dependences of Q'_I on V_{GB} . However, if V_{GB} is fixed and V_{CB} is varied instead, (3.4.28a) can be misleading. The reason is that the dependence of Q'_I on V_{CB} is very awkwardly predicted by it. It is hidden in Q'_M , V_{MB} , and n , each of which depends on V_{CB} in a complicated manner. On the other hand, (3.4.23) is ideal for such cases, since it makes explicit the exponential dependence of Q'_I on V_{CB} in a very simple manner.

3.4.4 Moderate Inversion

As seen in Fig. 3.2, moderate inversion is defined for a given V_{CB} by

$$V_{MB}(V_{CB}) \leq V_{GB} < V_{HB}(V_{CB}) \quad (3.4.30)$$

For a given V_{GB} value, one can numerically solve the implicit equation (3.2.7b) for ψ_s and substitute ψ_s into (3.2.6) to find Q'_I . Approximate explicit equations for ψ_s and the charges in terms of V_{GB} have been proposed.^{9-14,19,20} The developments of such expressions are lengthy and involve a number of approximations. Some representative results will be summarized in Sec. 3.5.3.

3.5 A “ V_{CB} CONTROL” POINT OF VIEW

3.5.1 Fundamentals

Our discussion of the three-terminal MOS structure so far was a natural extension of our discussion of the two-terminal structure. In both cases, we described in detail what happens when we increase the gate potential with respect to the substrate, V_{GB} . Thus, the gate potential was given the prominent role of the independent variable; the potential of terminal C with respect to the substrate, V_{CB} , was a parameter, as, for example, in Fig. 3.3. When we developed approximate expressions for the inversion layer charge in particular regions of inversion, often V_{CB} did not appear explicitly, but was rather hidden in the parameters in those equations, such as in V_{TB} in (3.4.11b) and in V_{MB} in (3.4.28). We will refer to the above view as “ V_{GB} control.”

There is another point of view, which we will call “ V_{CB} control,” in which we fix V_{GB} and observe what happens when we vary V_{CB} . Obviously, the two points of view should give equivalent results. However, each point of view adds something to our intuition and leads to a different set of approximations. In Chap. 4, it will be seen that the “ V_{CB} control” viewpoint leads to a different set of models for the MOS transistor,^{17,19,13} which have found their place along the classical “ V_{GB} control” models. It should be clear that “ x control,” as used here, does *not* imply any prominence of x in controlling the mechanisms in the three-terminal MOS structure, but rather refers to the way in which we choose to *describe* those mechanisms.

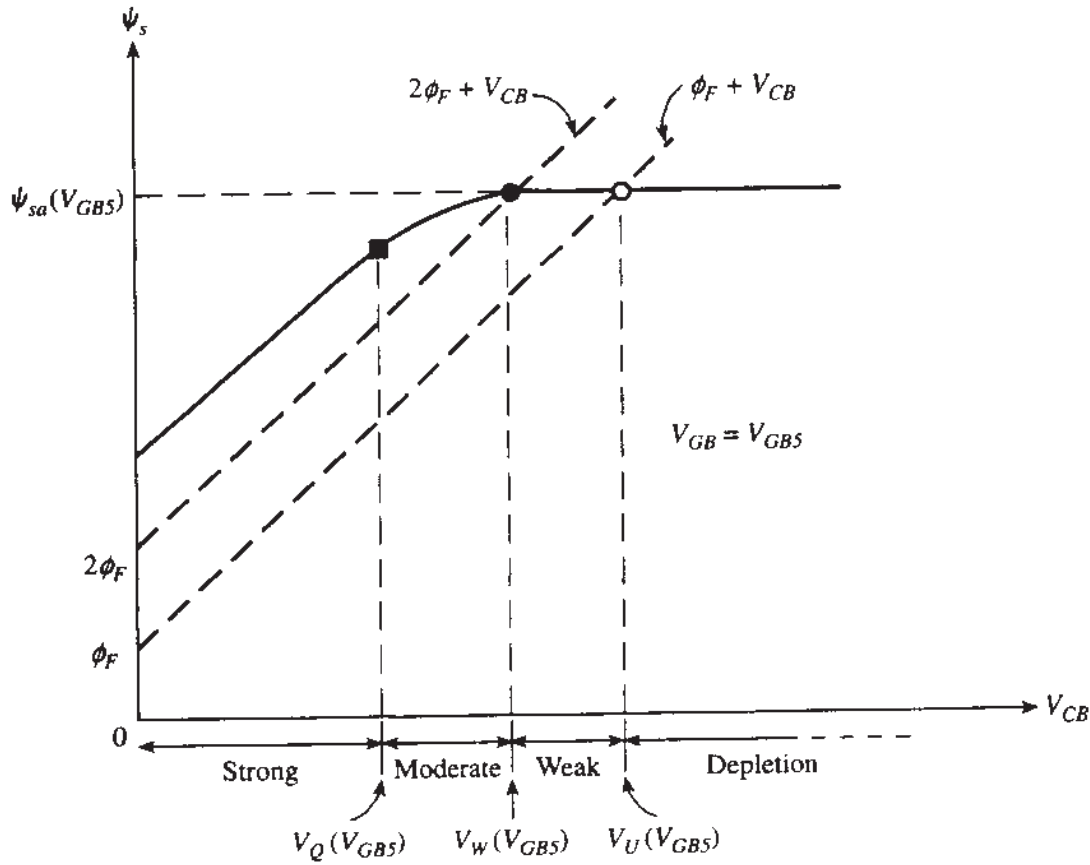


FIGURE 3.9
Surface potential vs. V_{CB} for a constant value of V_{GB} .

In order to make clear the connection between the two points of view, let us begin our discussion by returning to Fig. 3.3.

Let V_{GB} be *constant* and equal to the value V_{GB5} indicated on Fig. 3.3 (the subscript 5 is used in connection with a forthcoming discussion). As V_{CB} is raised, the level of inversion becomes lighter. For the values $V_{CB} = V_{CB1}$, V_{CB2} , and V_{CB3} the structure is correspondingly in strong, moderate, and weak inversion. Further increases in V_{CB} will leave the surface potential value practically unaffected at the value $\psi_{sa}(V_{GB5})$, as seen for example in going from V_{CB3} to V_{CB4} .

The above observations can clearly be displayed by plotting ψ_s vs. V_{CB} , with $V_{GB} = V_{GB5}$, using (3.2.7b). This is shown in Fig. 3.9. The quantities $V_Q(V_{GB5})$, $V_W(V_{GB5})$, and $V_U(V_{GB5})$ marked on the horizontal axis give the limits between strong and moderate inversion, between moderate and weak inversion, and between weak inversion and depletion, respectively, in terms of V_{CB} . The limit between moderate and weak inversion is seen to occur when $\psi_s \approx 2\phi_F + V_{CB}$, as expected from our earlier discussion. Similarly, the limit between weak inversion and depletion occurs when V_{CB} is increased to the point that $\psi_s = \phi_F + V_{CB}$ holds. For large V_{CB} , the inversion layer disappears and ψ_s flattens out at the value ψ_{sa} , which is given by (3.2.16) and *depends only on* V_{GB} . Note that as V_{CB} is increased, we go from strong, to moderate, to weak inversion. Thus, these regions are encountered in *opposite* order from the case where V_{GB} is increased (Fig. 3.3). This reflects the opposite roles of V_{CB} and V_{GB} in controlling the level of inversion.

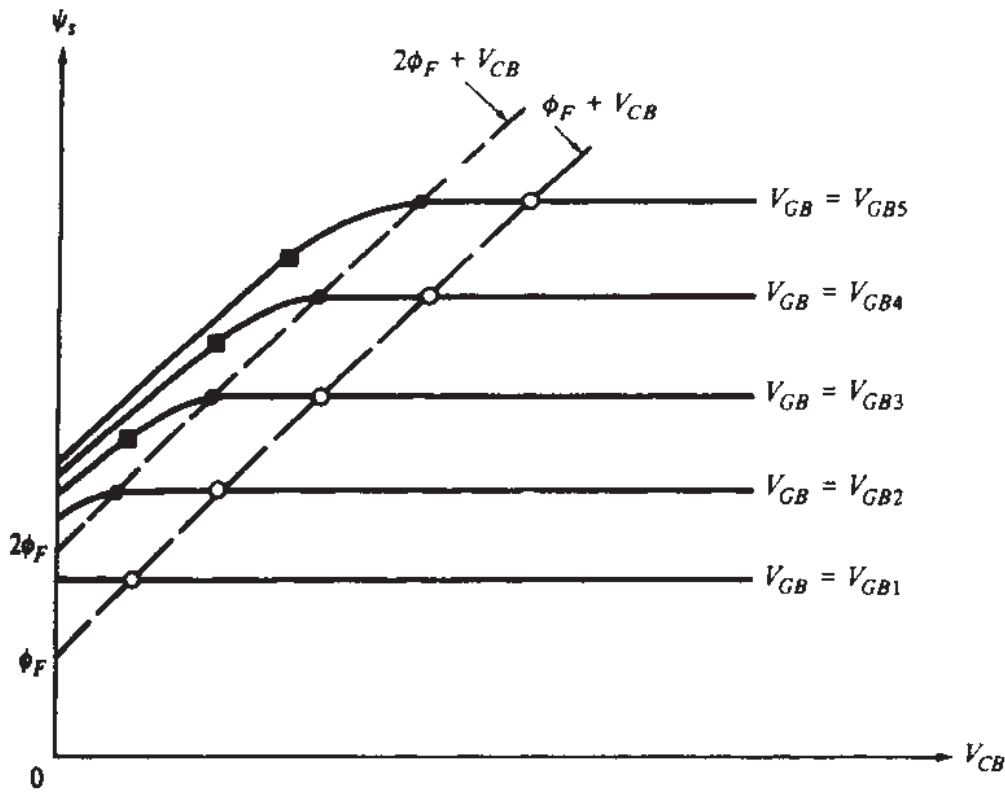


FIGURE 3.10

Surface potential vs. V_{CB} with gate-substrate voltage as a parameter. For each curve a full square denotes the limit between strong and moderate inversion; a full circle, the limit between moderate and weak inversion; and an open circle, the limit between weak inversion and depletion.

Adding curves for other values of V_{GB} produces the plot in Fig. 3.10. The two curves below the top one behave in a similar manner as above, only now V_{GB} is smaller ($V_{GB3} < V_{GB4} < V_{GB5}$ is assumed), and thus smaller values of V_{CB} are needed to reduce the level of inversion to a given point. The value $V_{GB2} (< V_{GB3})$ happens to be low, so even with $V_{CB} = 0$, the surface is only in moderate inversion. Increasing V_{CB} above zero can then only drive the structure into weak inversion and eventually into depletion. Finally, $V_{GB1} (< V_{GB2})$ is so low that the device is only in weak inversion when $V_{CB} = 0$. Increasing V_{CB} above 0 then will eventually drive the device into depletion.

Our analysis above would seem to predict that ψ_s just to the right of the n^+ region in Fig. 3.1 could be very different from the potential on the n^+ region itself (which, from Sec. 1.5, is $V_{CB} + \phi_{bi}$), depending on V_{CB} (Fig. 3.9). This would seem to indicate a discontinuity in the electrostatic potential as the n^+p boundary is crossed. This apparent problem is, of course, caused by our neglecting the "transition" region extending on both sides of the boundary. In this transition region the electric field is not vertical, and two-dimensional analysis becomes necessary. Such analysis can be carried out numerically and it does, indeed, predict that the potential changes from its value in the n^+ region to the value ψ_s , over a transition region the length of which is roughly equal to the depth of the depletion region under the n^+ region. The external source V_{CB} still "communicates" with the surface through the transition region, and the effect of V_{CB} to the right of the transition region can be expected to be as has been

described.† Unfortunately, two-dimensional analysis does not yield manageable analytical results.

The reader is urged to qualitatively produce Fig. 3.10 from Fig. 3.3, and vice versa. This is a good way to establish the connection between the “ V_{GB} control” and “ V_{CB} control” points of view.

Expressions for the bounds V_Q , V_W , and V_U between the various regions in terms of V_{CB} (see, for example, Fig. 3.9) can be developed by finding the values of V_{CB} needed to make the corresponding V_{GB} limits (in the second row of Table 3.1) equal to the given value of V_{GB} . Consider, for example, V_W . This quantity is the value of V_{CB} that brings the structure to the boundary between moderate and weak inversion. This means that if V_{CB} in (3.4.1) is set equal to V_W , the resulting value of V_{MB} must be equal to the given V_{GB} . In other words:

$$V_{FB} + 2\phi_F + V_W + \gamma\sqrt{2\phi_F + V_W} = V_{GB} \quad (3.5.1)$$

Solving this for V_W gives:

$$V_W = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right)^2 - 2\phi_F \quad (3.5.2)$$

This quantity is an increasing function of V_{GB} . This is because, if the structure is at the boundary between weak and moderate inversion and V_{GB} is subsequently increased, inversion will become heavier. To reduce inversion to its previous level, a larger V_{CB} will be needed, and therefore V_W will increase.

The expressions for V_W and for the other two limit points are summarized in the last row of Table 3.1.

The reader is reminded that V_{CB} is assumed non-negative, in order for the junction formed by the n^+ region and the substrate not to become forward-biased. In some cases, the expressions in the last row of Table 3.1 can result in negative or even imaginary values. This indicates that there are no positive (or zero) values of V_{CB} , which will bring the structure to the desired point. For example, consider the bottom curve in Fig. 3.10. In this case, V_{GB} is so low, that even with $V_{CB} = 0$ the structure is only in weak inversion. Raising V_{CB} above 0 can only drive the structure into depletion, as already remarked. Thus, only V_U will come out to have a real, positive value, which indicates that neither strong nor moderate inversion is possible for the given low value of V_{GB} .

In Fig. 3.11 we show how the behavior of ψ_s vs. V_{CB} , discussed above, affects the charges per unit area, Q'_i and Q'_b . As before, we assume that V_{GB} is fixed. For a

†Readers familiar with the use of quasi-Fermi levels (Appendix A) in pn junctions will recall that the electron quasi-Fermi level is taken practically constant from the n region through the depletion region.² In our case, the electron quasi-Fermi level (measured from the hole quasi-Fermi level in the bulk) corresponds to the potential V_{CB} (Appendix H).

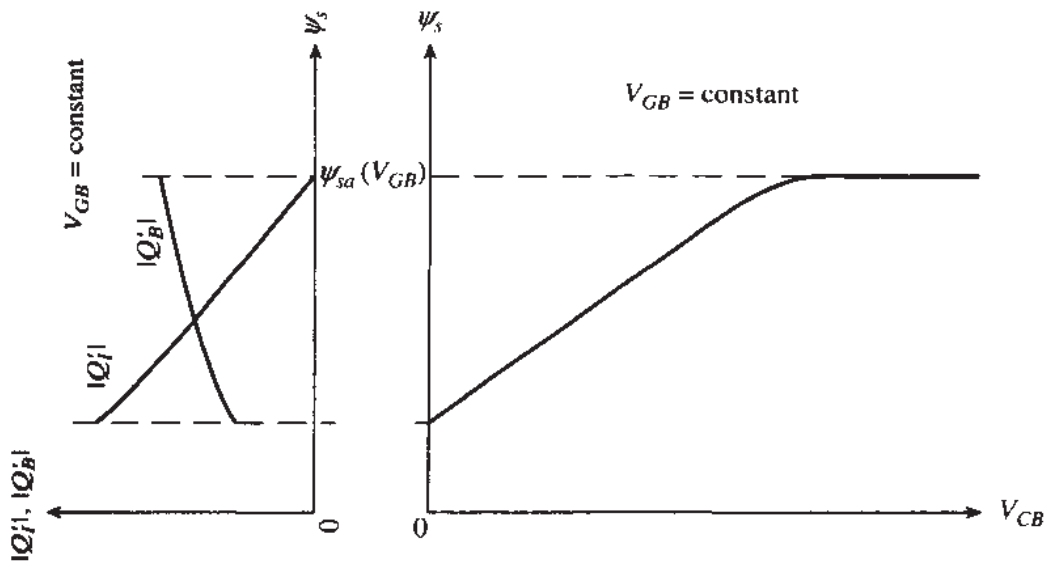


FIGURE 3.11

Surface potential vs. V_{CB} (right plot) and magnitudes of charges per unit area vs. surface potential (left plot).

given V_{CB} , ψ_s can be found as above, and is shown on the right. The resulting value of ψ_s determines Q'_I and Q'_B from (3.2.8b) and (3.2.5), respectively; this is shown in the left part of the figure. These plots are not shown for values of ψ_s larger than $\psi_{sa}(V_{GB})$, because ψ_s cannot attain such values; as seen on the right, ψ_s tends asymptotically to $\psi_{sa}(V_{GB})$, as V_{CB} is increased. Thus, Q'_B is expected to tend to a constant, maximum value and Q'_I is expected to tend to 0, as V_{CB} is raised.

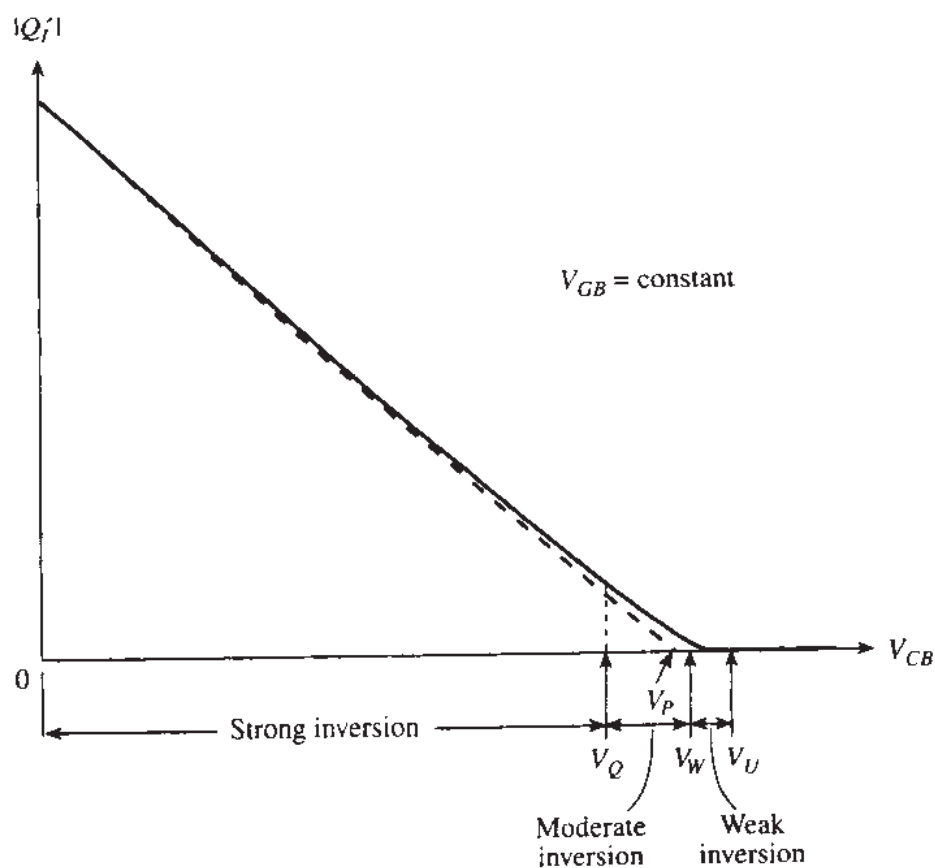
The behavior of $|Q'_I|$ vs. V_{CB} is shown in Fig. 3.12. This plot can be obtained from the parametric representation comprised of (3.2.7b) and (3.2.8b). The various regions of inversion are marked along the horizontal axis. As V_{CB} is raised past V_w , the structure leaves moderate inversion and enters weak inversion; the charge tends to 0 as V_{CB} is increased further. The broken line and the quantity V_p are related to approximations, and will be discussed in the following subsection. We emphasize that the plot in Fig. 3.12 is for a given V_{GB} ; if V_{GB} changes, the plot will also change, as shown in Fig. 3.13.

3.5.2 "Pinchoff" Voltage

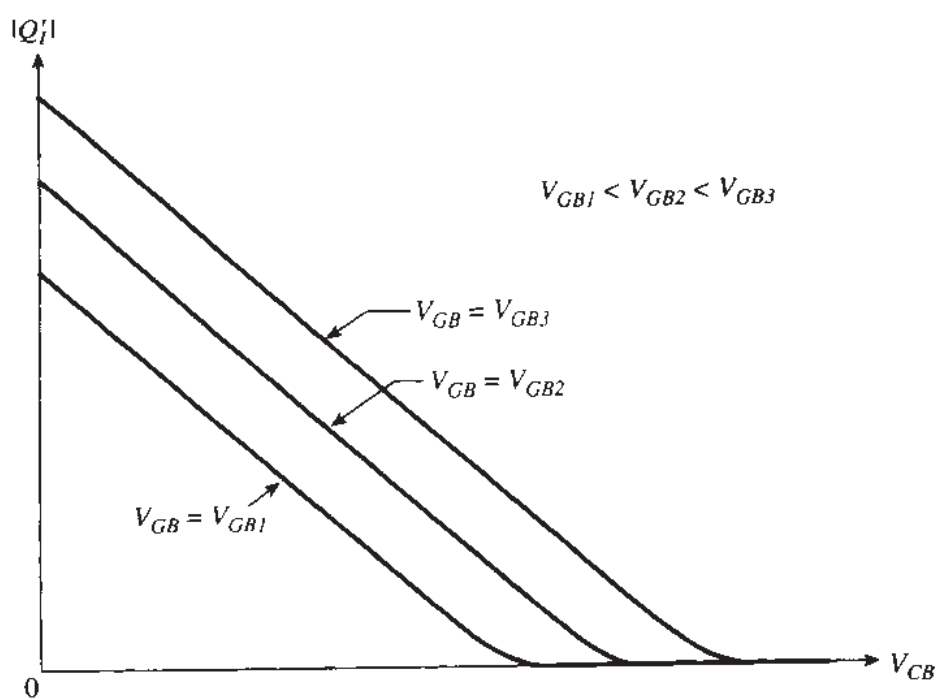
Consider the accurate plot of the inversion charge magnitude vs. V_{CB} , shown by the solid line in Fig. 3.12. In *strong* inversion, this charge can be approximated by (3.4.11b) with (3.4.12c) as

$$Q'_I = -C'_{ox} (V_{GB} - V_{FB} - \phi_0 - V_{CB} - \gamma \sqrt{\phi_0 + V_{CB}}) \quad (3.5.3)$$

The magnitude of this strong-inversion approximation is shown by the broken line in Fig. 3.12. As seen, it is close to a straight line, although it is *not* exactly a straight


FIGURE 3.12

Inversion layer charge magnitude vs. V_{CB} , for a fixed gate-substrate voltage. Solid line: charge sheet model; broken line: strong inversion approximation of (3.5.3).


FIGURE 3.13

Inversion layer charge magnitude vs. V_{CB} , for different gate-substrate voltages.

line. This plot crosses the horizontal axis at some V_{CB} value, shown as V_P in Fig. 3.12, which can be found by setting (3.5.3) equal to 0 and solving for V_{CB} . The result is

$$V_P = \left[-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right]^2 - \phi_0 \quad (3.5.4)$$

The quantity V_P is sometimes referred to as the “pinchoff” voltage. This name originates in the old days, when only strong inversion was important (or known), and the inversion layer was assumed to be “pinched off” (its charge was assumed to decrease to zero) at $V_{CB} = V_P$. As seen by the solid line in Fig. 3.12, however, at that point the structure is in moderate inversion, where the inversion charge has a significant value, and nothing is “pinched off.” Thus, we can say V_P is the value of V_{CB} that would have made Q'_I equal to zero, if strong inversion theory were valid for arbitrarily small $|Q'_I|$.

Another interpretation of V_P is as follows. Q'_I can be written using (3.4.11b) and (3.4.12c), repeated for convenience below:

$$Q'_I = -C'_{ox} [V_{GB} - V_{TB}(V_{CB})] \quad (3.5.5)$$

$$V_{TB}(V_{CB}) = V_{FB} + \phi_0 + V_{CB} + \gamma \sqrt{\phi_0 + V_{CB}} \quad (3.5.6)$$

Since V_P is the value of V_{CB} which makes the strong inversion charge zero, it is easy to see that V_P is the value of V_{CB} which makes $V_{TB}(V_{CB})$ equal to the applied gate-substrate voltage, V_{GB} . In other words:

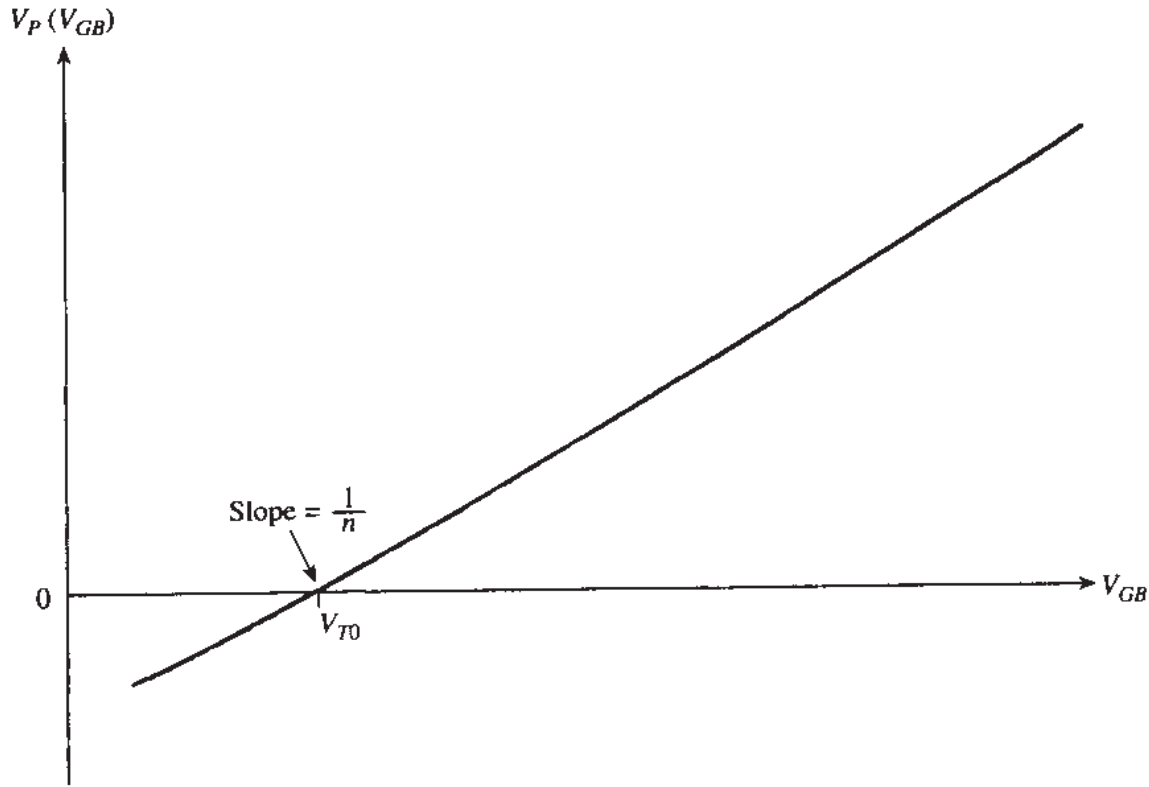
$$V_P = V_{CB} \Big|_{V_{TB} = V_{GB}} \quad (3.5.7)$$

Equation (3.5.4) can be written, using (3.4.16) to eliminate V_{FB} , as follows

$$V_P = V_{GB} - V_{T0} - \gamma \left[\sqrt{V_{GB} - V_{T0} + \left(\sqrt{\phi_0} + \frac{\gamma}{2} \right)^2} - \left(\sqrt{\phi_0} + \frac{\gamma}{2} \right) \right] \quad (3.5.8)$$

A plot of V_P versus V_{GB} is shown in Fig. 3.14. The fact that the plot crosses the horizontal axis at $V_{GB} = V_{T0}$ is consistent with the interpretation of V_P in (3.5.7). The slope of the plot will be discussed in the following subsection.

†As already mentioned, in some treatments ϕ_0 is approximated by $2\phi_F$. In such cases, V_P becomes equal to V_w , as seen from (3.5.4) and (3.5.2).

**FIGURE 3.14**

The parameter V_P as a function of the gate-substrate voltage.

3.5.3 Expressions in Terms of the “Pinchoff” Voltage

A comparison of (3.5.4) to (3.2.16) shows that:

$$\psi_{sa} = V_P + \phi_0 \quad (3.5.9)$$

The inverse of the slope of ψ_{sa} vs. V_{GB} was denoted in (3.2.17) by n . Since $V_P(V_{GB})$ and $\psi_{sa}(V_{GB})$ differ by a constant, they have the same slope, so we can write

$$n = \left(\frac{dV_P}{dV_{GB}} \right)^{-1} \quad (3.5.10)$$

In some treatments,¹⁹ this is used as the definition of n . The value of n in terms of V_P can be found from (3.2.18) with (3.5.9):

$$n = 1 + \frac{\gamma}{2\sqrt{\phi_0 + V_P(V_{GB})}} \quad (3.5.11)$$

Note that n has a well-defined value for a given V_{GB} , independent of V_{CB} .

We now develop a simple approximation¹⁹ for V_P . We see from (3.5.10) that the slope of $V_P(V_{GB})$ (Fig. 3.14) is equal to $1/n$. Thus, if the plot is approximated by a straight line in Fig. 3.14, we can write:¹⁹

$$V_P \approx \frac{V_{GB} - V_{T0}}{n} \quad (3.5.12)$$

where, strictly speaking, n should be evaluated at the point $V_{GB} = V_{T0}$. However, the approximation in (3.5.12) is used¹⁹ with $n = n(V_{GB})$, as evaluated from (3.5.11) with (3.5.4). The slowly decreasing $n(V_{GB})$ (see Fig. 3.4) imparts a curvature to the plot of V_P from (3.5.12) and approximates the curve in Fig. 3.14 rather closely.

We now consider expressions for Q'_I as a function of V_{CB} and V_P . These are used in conjunction with several models based on the " V_{CB} control" point of view.^{17,19,13}

STRONG INVERSION. The strong-inversion approximation for Q'_I was given by (3.5.3), and its magnitude was shown by the broken line in Fig. 3.12. From (3.5.3) we have:

$$\left. \frac{dQ'_I}{dV_{CB}} \right|_{V_{CB}=V_P} = C'_{ox} \left(1 + \frac{\gamma}{2\sqrt{\phi_0 + V_{CB}}} \right) \bigg|_{V_{CB}=V_P} \quad (3.5.13a)$$

which, if we use (3.5.11) gives

$$\left. \frac{dQ'_I}{dV_{CB}} \right|_{V_{CB}=V_P} = C'_{ox} n \quad (3.5.13b)$$

Therefore, the strong-inversion approximation can be approximated by a first-order expansion around $V_{CB} = V_P$ as follows¹⁹:

$$Q'_I \approx \left(\left. \frac{dQ'_I}{dV_{CB}} \right|_{V_{CB}=V_P} \right) (V_{CB} - V_P) \quad (3.5.14a)$$

$$Q'_I \approx -nC'_{ox}(V_P - V_{CB}) \quad (3.5.14b)$$

One may question the point around which linearization was performed above, as opposed to, say, the point $V_{CB} = 0$ in Fig. 3.12. However, it is easy to prove that the latter would have resulted in a large error near $V_{CB} \approx V_Q$.

WEAK INVERSION. In the approximation (3.5.14b), Q'_I in strong inversion is a function of $V_P - V_{CB}$. It is easy to produce a weak inversion equation which also depends on the same quantity. Using ψ_{sa} from (3.5.9) in (3.4.23), we obtain

$$Q'_I = -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\phi_0 + V_P}} \phi_t e^{(\phi_0 - 2\phi_F)/\phi_t} e^{(V_P - V_{CB})/\phi_t} \quad (3.5.15)$$

where, again, V_P and n depend on V_{GB} , as already discussed.

MODERATE INVERSION. Several approximate expressions have been suggested in the literature for the surface potential and the inversion layer charge in moderate inversion.^{9-14,20} The development of such expressions is either empirical or very lengthy, or a mixture of the two.

Although we will not present such development here, we will show some representative results,[†] which parallel the ones we gave for the two-terminal structure in Sec. 2.5.4. Readers interested in the justification of such results can consult the references provided.

As expected from Fig. 3.7, for the surface potential in moderate inversion one can use

$$\psi_s = 2\phi_F + V_{CB} + \Delta\phi \quad (3.5.16)$$

As seen in the same figure, for success in moderate inversion $\Delta\phi$ must be made a function of V_{GB} . The following approximation has been suggested¹³:

$$\Delta\phi = \frac{2\phi_t}{n} \ln \left(1 + \frac{V_P - V_{CB} + \phi_0 - 2\phi_F}{2\phi_t} \right) \quad (3.5.17)$$

This expression is consistent with (2.5.47), reducing to the latter for $V_{CB} = 0$. To make possible continuity from moderate to strong inversion, (3.5.16) and (3.5.17) can be used also in strong inversion. In weak inversion, (3.4.22) should be used, as in that region (3.5.16) and (3.5.17) produce meaningless results.

It has been suggested that the inversion charge in moderate inversion can be approximated by¹³:

$$Q'_i = - \frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\phi_0 + V_P}} \phi_t - nC'_{ox}(V_P - V_{CB} + \phi_0 - 2\phi_F - \Delta\phi) \quad (3.5.18)$$

which is consistent with (2.5.48). At the boundary between moderate and weak inversion, the above equation reduces to the value given by the weak inversion relation (3.5.15), which should then take over. Equation (3.5.18) can be used also in strong inversion in order to maintain continuity.¹³

While the expressions discussed in this section provide explicit relations for individual regions of inversion, the use of a single equation in all regions of inversion is, of course, the ultimate goal of modeling and avoids "if" statements in a computer implementation. Thus, if the extra calculation time involved in using (3.2.7b) can be afforded, the use of that equation and (3.2.6) is certainly appealing.

REFERENCES

1. H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator) semiconductor transistors," *Solid-State Electronics*, vol. 9, pp. 927-937, 1966.
2. S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.
3. I. R. M. Mansour, "On the modeling of MOS devices," *Proceedings of the Third International Symposium on Network Theory*, Yugoslavia, pp. 705-713, 1975.
4. I. R. M. Mansour, "Improved modeling of MOS devices," *Proceedings of the European Conference on Circuit Theory and Design*, Italy, 1976.

[†]Material set in smaller type can be skipped without loss of continuity.

5. G. Baccarani, M. Rudan, and G. Spadini, "Analytical i.g.f.e.t. model including drift and diffusion currents," *IEEE Journal of Solid-State and Electron Devices*, vol. 2, pp. 62–68, March 1978.
6. J. R. Brews, "A charge-sheet model of the MOSFET," *Solid-State Electronics*, vol. 21, pp. 345–355, 1978.
7. F. Van de Wiele, "A long-channel MOSFET model," *Solid-State Electronics*, vol. 22, pp. 991–997, 1979.
8. J. R. Brews, "Physics of the MOS transistor," chap. 1 in *Silicon Integrated Circuits, Part A*, D. Kahng (editor), Applied Solid State Science Series, Academic Press, New York, 1981.
9. L. L. Lewlyn and J. D. Meindl, "An IGFET inversion charge model for VLSI systems," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 434–440, February 1985.
10. M. A. Maher and C. A. Mead, "A physical charge-controlled model for MOS transistors," pp. 211–229, from P. Lesleben (editor), *Advanced Research VLSI*, Cambridge, Mass., The MIT Press, 1987.
11. C.-K. Park, C.-Y. Lee, B.-J. Moon, Y. H. Byun, and M. Shur, "A unified current-voltage model for long-channel nMOSFET's," *IEEE Transactions on Electron Devices*, vol. 38, pp. 399–406, February 1991.
12. B. Iñiguez and E. G. Moreno, "Explicit C_{∞} -continuous and general model for nMOSFETs," *Electronics Letters*, vol. 29, pp. 1036–1037, 27 May 1993.
13. A. I. A. Cunha, M. C. Schneider, and C. G.-Montoro, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electronics*, vol. 38, pp. 1945–1952, November 1995.
14. K. Lee, M. Shur, T. A. Fjeldly, and T. Ytterdal, *Semiconductor Device Modeling for VLSI*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
15. Y. Tsividis, "Moderate inversion in MOS devices," *Solid-State Electronics*, vol. 25, pp. 1099–1104, 1982; see also Erratum, *ibid.*, vol. 26, p. 823, 1983.
16. R. M. Swanson and J. D. Meindl, "Ion-implanted complementary MOS transistors in low-voltage circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-7, pp. 146–153, April 1972.
17. E. Vittoz and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operation," *IEEE Journal of Solid-State Circuits*, vol. SC-12, pp. 224–231, June 1977.
18. R. J. Van Overstraeten, G. J. Declercq, and P. A. Nuls, "Theory of the MOS transistor in weak inversion—new method to determine the number of surface states," *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 282–288, May 1975.
19. C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog Integrated Circuits and Signal Processing*, vol. 8, pp. 83–114, July 1995.
20. V. Altschul and Y. S.-Diamand, "Modeling of the MOSFET inversion charge and drain current in moderate inversion," *IEEE Transactions on Electron Devices*, vol. 37, pp. 1909–1915, August 1990.

PROBLEMS

- 3.1. (a) For a device with $N_A = 10^{17} \text{ cm}^{-3}$, $t_{\text{ox}} = 100 \text{ \AA}$, and $V_{FB} = -1 \text{ V}$, plot ψ_s vs. V_{GB} , with V_{GB} between 0.5 and 5 V and for $V_{CB} = 1, 2$, and 3 V. Show approximately the three regions of inversion on each curve. Assume that $V_{HB} \approx V_{MB} + 0.55 \text{ V}$.
 (b) For the same device plot ψ_s vs. V_{CB} , with V_{CB} between 0 and 5 V and for $V_{GB} = 1, 2$, and 3 V. Show approximately the regions of inversion on each curve.
- 3.2. Show how the V_{CB} - V_{GB} plane can be separated into five regions, corresponding to accumulation, depletion, weak inversion, moderate inversion, and strong inversion. (Hint: In accumulation, take into account the fact that the concentration of holes is practically unaffected by V_{CB} .)
- 3.3. For the device of Prob. 3.1:
 - (a) Find V_{LB} , V_{MB} , and V_{HB} for $V_{CB} = 0 \text{ V}$; assume $V_{HB} \approx V_{MB} + 0.55 \text{ V}$.

- (b) Plot $|Q'_I|$ vs. V_{GB} for V_{GB} between V_{LB} and $V_{HB} + 5$ V. Use (3.2.6) and (3.2.7b). Give the corresponding plot vs. V_{GC} .
- (c) Repeat using a logarithmic axis for $|Q'_I|$.
- (d) On each of the above plots, plot $|Q'_I|$ as given by (3.4.11b) in strong inversion and as given by (3.4.28b) in weak inversion. Comment on the accuracy obtainable.
- (e) Repeat (a) through (d) for $V_{CB} = 3$ V.
- 3.4. Plot $\ln |Q'_I|$ vs. ψ_s , and verify that this plot approaches straight line behavior in the weak and strong inversion regions. Use the process parameters of Prob. 2.6 and $V_{CB} = 1$ V. Determine and compare the slopes of the plot in these two regions.
- 3.5. It is often assumed that there is a point below which $Q'_I(V_{GB})$ is an exponential and above which it is a straight line. Attempt to find such a point on the plots of Prob. 3.3. Is this possible? How large an error is involved in such an approximation?
- 3.6. It has been seen in Sec. 3.4.2 that if ψ_s is pinned to a fixed value, Q'_I plots as a straight line vs. V_{GB} . Is such pinning a *necessary* condition for obtaining such a straight line? If yes, why? If not, find the necessary condition for such a behavior of ψ_s .
- 3.7. Show that a better approximation than (3.4.6) for the surface potential in *strong* inversion is:

$$\psi_{s, strong} = 2\phi_F + V_{CB} + \phi_i \ln \left\{ \frac{1}{\phi_i} \left[\frac{1}{\gamma^2} (V_{GB} - V_{FB} - V_{CB} - \phi_0)^2 - \phi_0 - V_{CB} \right] \right\}$$

[Hint: Use (3.4.6) as an initial guess and perform one iteration using (3.2.7b).] Compare the above expression to (3.4.6) and to (3.2.7b). What happens if we attempt to use the above expressions in moderate and weak inversion?

- 3.8. Equation (3.4.28b) was developed by using an approximation around $\psi_s = 2\phi_F + V'_{CB}$. Develop a similar expression using an expansion around $\psi_s = 1.5\phi_F + V'_{CB}$. How do the two expressions compare?
- 3.9. Show that for the structure of Fig. 3.1c, $C'_i < 0.1C'_b$ if the surface potential satisfies $\psi_s < 2\phi_F + V_{CB} - 2.3\phi_i$.
- 3.10. For a given V_{GB} , find *explicit* expressions for ϕ_M , ϕ_H , V_w , and V_Q using the procedure outlined at the end of Appendix G.
- 3.11. Mark the accurate limits of the regions of inversion on the curves of Prob. 3.3 using the definitions of Appendix G.
- 3.12. Prove (3.2.6). [Hint: Find the total semiconductor charge $Q'_C = Q'_I + Q'_B$ using basic electrostatics (Appendix B) by following a procedure similar to that of Prob. 2.12. From that find Q'_I by subtracting (3.2.5a).]
- 3.13. Verify the entries in Table 3.1.
- 3.14. Verify the entries in Table 3.2.

CHAPTER 4

THE FOUR- TERMINAL MOS TRANSISTOR

4.1 INTRODUCTION

The MOS transistor is obtained by adding one more terminal to the structure of Chap. 3, so that the inversion layer is contacted at two opposite ends. By applying a voltage between these ends, a current can be caused to flow in the layer. Since the density of carriers available for conduction depends on the gate potential, the latter can be used to either create or eliminate the inversion layer (i.e., turn the device “on” or “off”) for digital applications or to modulate its conduction in a continuous manner for analog applications.

The path from the conception of the basic MOS transistor principle to the demonstration of working devices is spread over a quarter century¹⁻⁴; see also Ref. 5. The basic theory of MOS transistor operation was developed in the early nineteen-sixties.⁶⁻⁹ Extensive treatments on the MOS transistor were published in the late sixties, including investigations of the role of diffusion current and the role of substrate charge.¹⁰⁻¹⁸ Our list of references to the work of the sixties is by no means complete, and many more references can be found in several early texts that helped spread the knowledge of the device.^{11,18-20} Starting in the sixties, the industry applied the newly acquired knowledge to the successful fabrication of digital ICs, and the need arose for efficient models for computer-aided design.²¹⁻²⁵ At the same time, the current in weakly inverted channels received considerable attention.²⁶⁻³⁹ Starting in the

seventies, attempts have been made to produce models valid for all combinations of external bias voltages by including drift and diffusion currents,^{40–47} but in a simpler manner than that of early attempts.¹⁰ The above references are only indicative of early work in the topics mentioned in this chapter; further references on such topics will be given along the way. Extensive work has also been done on mobility behavior, ion-implanted devices, short- and narrow-channel devices, noise, the modeling of charges and capacitances, high-frequency operation, and other topics, as well as on extensive models for computer-aided design, incorporating many different effects. Related references will be given as such topics are discussed in this and in subsequent chapters.

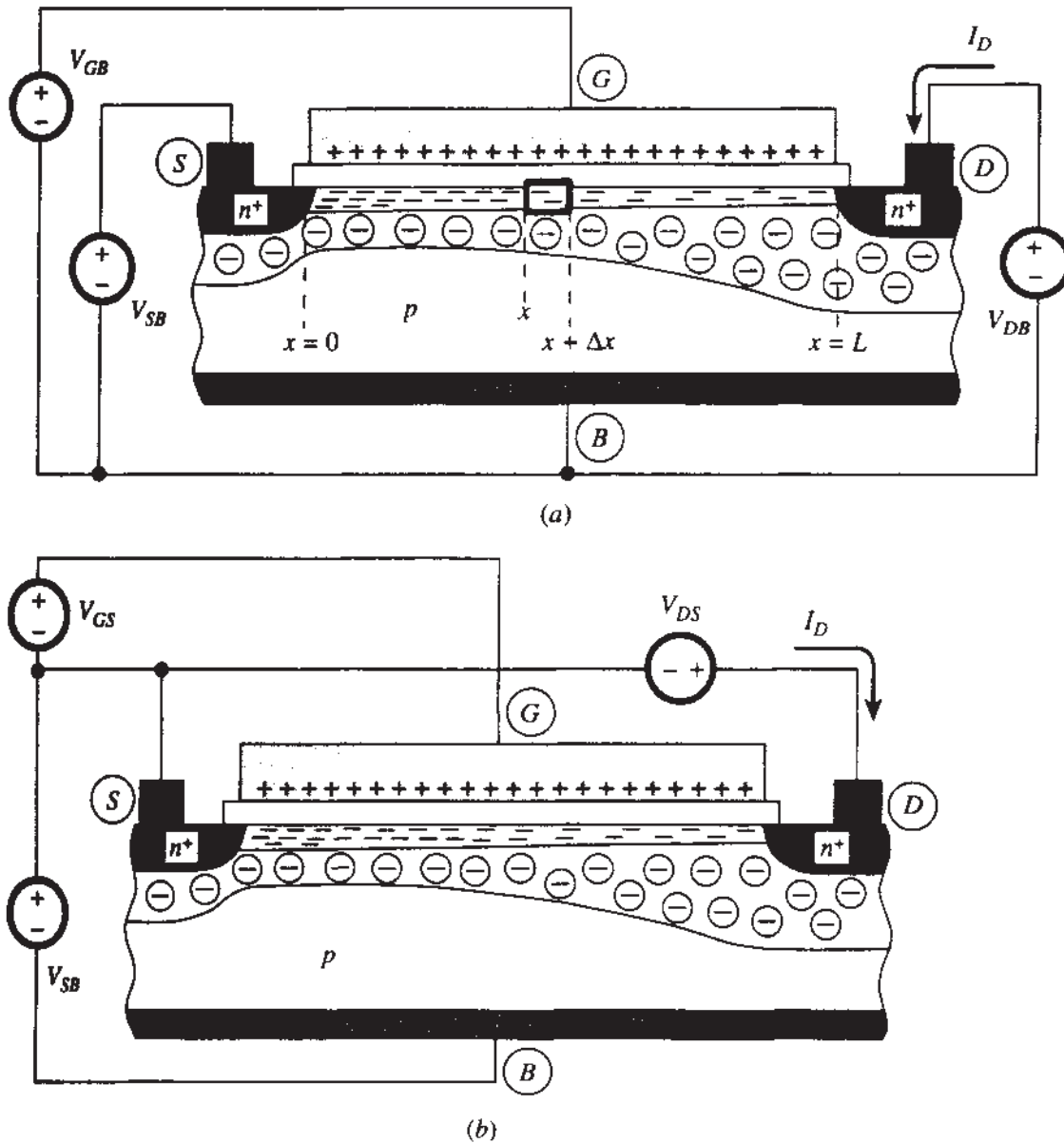
The basic structure of a MOS transistor and a brief overview of its operation and characteristics were given in Sec. 1.6. In this chapter, our goal will be to determine the drain current for any combination of dc terminal voltages. Throughout the chapter we will assume that the channel is sufficiently *long* and *wide*, so that edge effects are confined to a negligible part of it; this statement will become more quantitative later on. We will also assume that the substrate is *uniformly* doped. (The doping will be assumed to be *p* type unless indicated otherwise.) There have been many models proposed for this basic case, and many readers are likely to encounter several of them in the course of their work. In this chapter, we will introduce these models in a systematic way. Although each model is usually derived in the literature independently of the others, here we will place all of them in context, by showing how they can all be derived from one important model, and how they relate to that model and to each other. Readers who do not desire a very detailed treatment can skip certain sections without loss of continuity. We indicate so explicitly in the beginning of such sections, or set them in fine print.

An nMOS transistor with external dc voltages applied is shown in Fig. 4.1. In this figure, and all related figures in this book, we show only the part of the depletion region extending into the substrate. The extent of this region into the heavily doped n^+ regions is much narrower and is not shown for simplicity. (The complete picture of the depletion region would be qualitatively similar to that shown in Fig. 1.15.) In Fig. 4.1a we show a common substrate connection for the voltage sources, whereas in Fig. 4.1b we show a common source connection; both cases will prove useful in our ensuing discussion. We will be able to apply at the source end of the channel the results we developed for the structure of Fig. 3.1c or d, with $V_{CB} = V_{SB}$. Similarly, at the drain end, we will be able to apply those results with $V_{CB} = V_{DB}$. It should be emphasized that *normal operation of a MOS transistor requires that both pn junctions be reverse-biased*. Thus, for an nMOS transistor it will be implicitly assumed that†

$$V_{SB} \geq 0 \quad (4.1.1)$$

$$V_{DB} \geq 0 \quad (4.1.2)$$

†In MOS transistor models used in circuit simulation, diodes must be included between source and substrate, and between drain and substrate. This is in order to take care of cases where (4.1.1) and (4.1.2) are not satisfied in the circuit, and also to model the leakage currents of the reverse-biased n^+p junctions. The n^+p junction currents are assumed to be modeled separately in this way, and are not considered here.

**FIGURE 4.1**

An n -channel MOS transistor. (a) Terminal voltages referred to the substrate; (b) terminal voltages referred to the source.

We assume that the external dc voltages have been applied for a long time, so that all charges have reached steady-state values.

If $V_{DB} = V_{SB}$ in Fig. 4.1a, we have a situation similar to that in Fig. 3.1c, and the electric field in the semiconductor is perpendicular to the surface (except at points too close to the n^+ regions). If now $V_{DB} \neq V_{SB}$, there will be a nonzero component of the electric field in the horizontal direction. *The horizontal field component will be assumed to be much smaller than the vertical component.* Thus the field direction is assumed to be practically the same as in Fig. 3.1c, and the equations developed in reference to that figure will be appropriately extended for use here. (This approximation is referred to as the *gradual channel approximation* and can be verified with

two-dimensional numerical computations; however, we will also encounter cases later where this approximation fails.)

In this chapter, we will view the oxide and the depletion region under the channel as perfectly insulating layers. This means that for the gate current I_G and for the substrate current I_B we will assume:

$$I_G = 0 \quad (4.1.3)$$

$$I_B = 0 \quad (4.1.4)$$

It will be seen in Chap. 6 that this assumption can be violated if the horizontal component of the electric field becomes so large that it can produce very energetic electrons; this can cause a nonzero I_G and, most importantly, a nonzero I_B which must be added to the channel current, to give the total drain current. Also, at high temperatures, the drain-body junction can have a significant leakage current, which must also be included in the drain current. Thus, in general, the drain current I_D can be different from the channel current, or *drain-to-source current*, I_{DS} . In this chapter, unless noted otherwise, we will assume that the above effects are negligible, and thus

$$I_D = I_{DS} \quad (4.1.5)$$

Whereas in Fig. 3.1c the various charges were assumed to be uniformly distributed along the x direction (except for points too close to the n^+ region), this will, in general, not be the case in Fig. 4.1, because potentials will vary with x . Thus, charges per unit area should be defined *locally*. For example, consider a small area element ΔA of the channel (as seen from above in Fig. 4.1) centered around x , and the inversion layer charge ΔQ_I contained in it. The inversion layer charge per unit area Q'_I at point x is defined as the limit of $\Delta Q_I/\Delta A$, as ΔA approaches 0; similarly for Q'_B and Q'_G . In other words, we define:

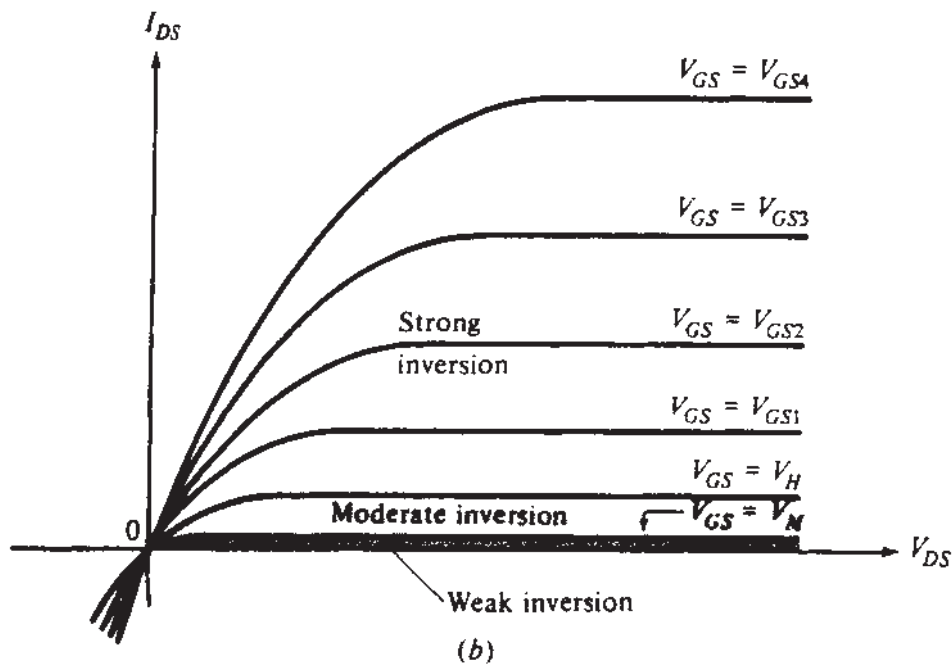
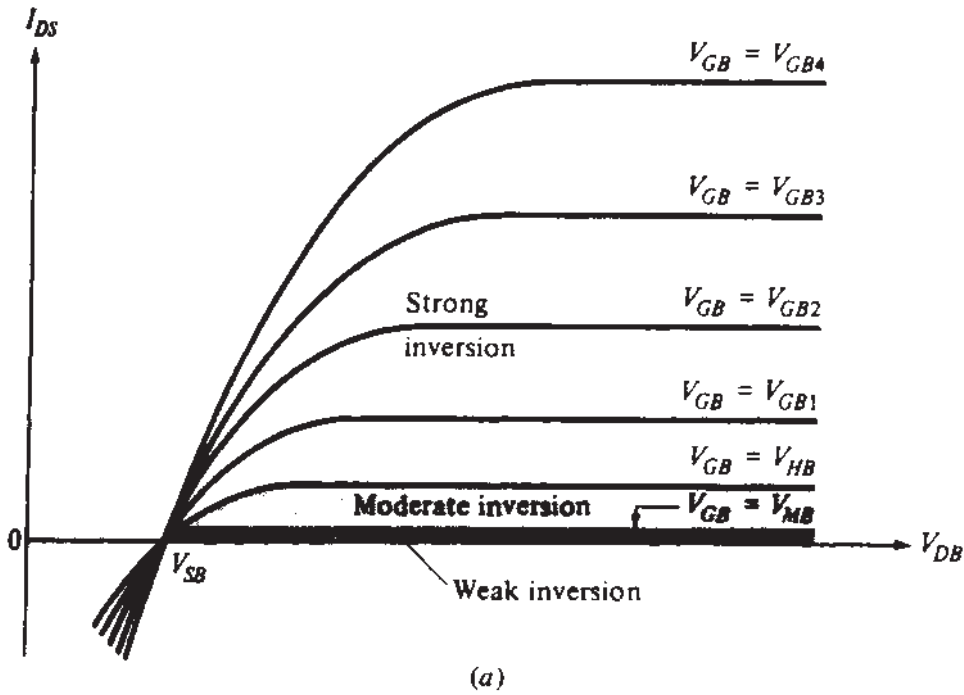
$$Q'_I = \frac{dQ_I}{dA} \quad (4.1.6)$$

$$Q'_B = \frac{dQ_B}{dA} \quad (4.1.7)$$

$$Q'_G = \frac{dQ_G}{dA} \quad (4.1.8)$$

4.2 TRANSISTOR REGIONS OF OPERATION

Typical sets of dc current-voltage characteristics for an NMOS transistor are shown in Fig. 4.2a and b, corresponding to Fig. 4.1a and b, respectively. If the *logarithm* of the current is plotted on the vertical axis, plots of the type shown in Fig. 4.3 are obtained. This type of plot is more fair to the moderate and weak inversion regions, which are very important in low-voltage, low-power applications.

**FIGURE 4.2**

Current-voltage characteristics corresponding to Fig. 4.1a and b. In both cases V_{SB} is assumed fixed, and $V_{GBi} = V_{GSi} + V_{SB}$.

On each curve in Figs. 4.2 and 4.3, the initial, sloped part is termed *nonsaturation*. With sufficiently high drain potential, the curve flattens out and we have *saturation*.† The bound between nonsaturation and saturation will be discussed later.

†In some of the literature, the names “nonsaturation” and “saturation” are used in conjunction with strong inversion only. Since the same qualitative behavior is observed also in moderate and weak inversion, we will be using the above names in any region.

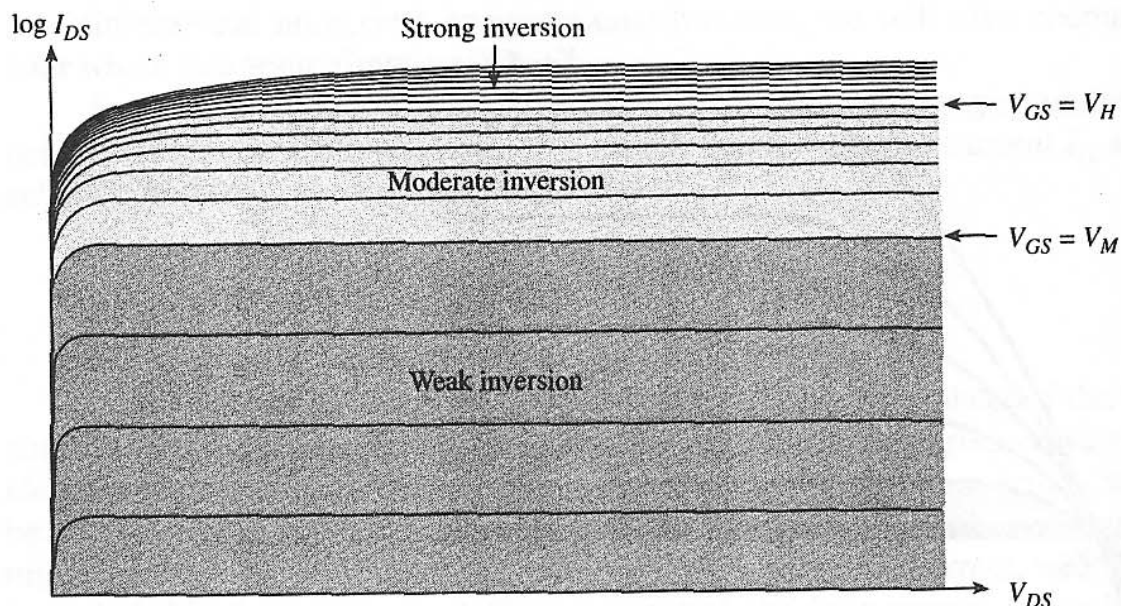


FIGURE 4.3

Logarithm of drain-source current vs. drain-source voltage, with gate-source voltage as a parameter.

TABLE 4.1
Transistor regions of operation

| Region | Channel condition |
|--------------------|---|
| Strong inversion | The more heavily† inverted channel end is in strong inversion |
| Moderate inversion | The more heavily† inverted channel end is in moderate inversion |
| Weak inversion | The more heavily† inverted channel end is in weak inversion |

†If both ends are equally inverted, either end can be considered.

In general, the drain current depends on the terminal voltages, in a complicated manner. However, for certain combinations of terminal voltages, simplifications become possible and relatively simple expressions can be developed for the current. Anticipating this, we will define transistor regions of operation as shown in Table 4.1. These definitions are consistent with long-established practice, and the rationale behind them will be seen in subsequent sections. As a mnemonic aid it should be noted that:

The name of a region coincides with the level of inversion at the more heavily inverted channel end.

For convenience in drawing certain figures (like Fig. 4.1) or providing certain plots, we will be assuming that the more heavily inverted channel end is the one next to the *source*, unless noted otherwise. This does not restrict generality, since the source and drain terminals in Fig. 1.20 are equivalent.

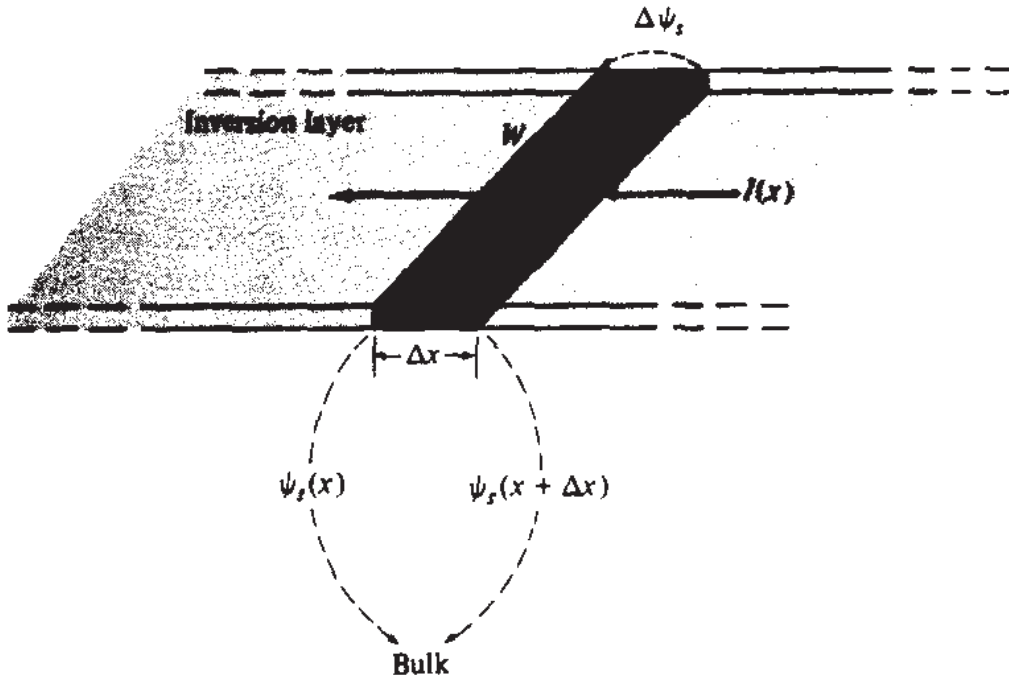


FIGURE 4.4
Small element of the inversion layer in the device of Fig. 4.1.

4.3 GENERAL CHARGE SHEET MODELS

4.3.1 Complete Charge Sheet Model

In this section, we derive expressions for the drain current, valid in all regions of inversion. The term *general* in the title refers to this universal validity. The term *charge sheet* refers to the basic assumption in the model, i.e., that the inversion layer is of infinitesimal thickness.[†] We note that we have already made this assumption in Chaps. 2 and 3. We will use expressions from those chapters in our present development. The model we will present has been derived in several references,^{40,41,43–46} but our derivation will be simpler.

A key to the generality of the results we are about to develop is the observation that the current in the channel can be caused by both drift and diffusion (Sec. 1.3). Thus, let x be the horizontal position in the channel, measured from the source end. If the inversion layer current at x is denoted by $I(x)$, we will have

$$I(x) = I_{\text{drift}}(x) + I_{\text{diff}}(x) \quad (4.3.1)$$

To write an expression for the drift component, consider a small element in the inversion layer between x and $x + \Delta x$ in Fig. 4.1a, as shown magnified in Fig. 4.4. The potential difference across this element is $\Delta\psi_s(x) = \psi_s(x + \Delta x) - \psi_s(x)$. Comparing this figure to Fig. 1.5 [and assuming that the electron velocity is proportional to

[†]Note, nevertheless, that when drawing a figure it is convenient to use a finite thickness for the inversion layer.

the small horizontal electric field, analogous to (1.3.8)], it is obvious that we can use (1.3.15) with V replaced by $\Delta\psi_s(x)$, a replaced by Δx , b by W , and $|Q'|$ by $-Q'_i$, where Q'_i is the (negative) inversion layer charge per unit area at x . Also, μ_B must be replaced with a *smaller* mobility value because electrons move with difficulty parallel to the surface (the semiconductor-oxide interface), being pulled toward it by the vertical field. The corresponding mobility will be called the *surface mobility* and will be denoted by μ . (This quantity will be considered in more detail in Sec. 4.10; for now it can be assumed that it has roughly half the value of μ_B .) We will thus have

$$I_{\text{drift}}(x) = \mu(-Q'_i) \frac{W}{\Delta x} \Delta\psi_s(x) \quad (4.3.2)$$

which, allowing Δx to approach zero, becomes

$$I_{\text{drift}}(x) = \mu W(-Q'_i) \frac{d\psi_s}{dx} \quad (4.3.3)$$

The diffusion current component can be obtained as in (1.3.20):

$$I_{\text{diff}}(x) = \mu W \phi_i \frac{dQ'_i}{dx} \quad (4.3.4)$$

We should note here that, if the inversion layer were assumed to be of finite thickness and the electron concentration in it varied with depth, (4.3.3) and (4.3.4) would still be valid, assuming a "laminar" electron flow. This is owing to the generality of (1.3.15) and (1.3.20), discussed in the corresponding footnotes in Sec. 1.3. However, this point is not of much importance here, where it is assumed that all inversion layer charge is concentrated within an infinitesimal depth from the "surface."

In dc steady state, which is the only case discussed in this chapter, the total current in the channel must be the same for all x and equal to the drain-to-source current (which, under our assumptions of Sec. 4.1, is equal to the drain current). Using this fact and (4.3.3) and (4.3.4) in (4.3.1), we obtain

$$I_{DS} = \mu W(-Q'_i) \frac{d\psi_s}{dx} + \mu W \phi_i \frac{dQ'_i}{dx} \quad (4.3.5)$$

Let the surface potential at the source end of the channel ($x = 0$) be denoted by ψ_{s0} , and Q'_i there by Q'_{i0} . Let the corresponding quantities at the drain end of the channel ($x = L$) be denoted by ψ_{sL} and Q'_{iL} . Integrating (4.3.5) from $x = 0$ to $x = L$ we obtain

$$\int_0^L I_{DS} dx = W \int_{\psi_{s0}}^{\psi_{sL}} \mu(-Q'_i) d\psi_s + W \phi_i \int_{Q'_{i0}}^{Q'_{iL}} \mu dQ'_i \quad (4.3.6)$$

Since I_{DS} is independent of x it can be moved outside the integral. Thus the left-hand side is equal to $I_{DS}L$ and we have

$$I_{DS} = \frac{W}{L} \left[\int_{\psi_{s0}}^{\psi_{sL}} \mu(-Q'_i) d\psi_s + \phi_i \int_{Q'_{i0}}^{Q'_{iL}} \mu dQ'_i \right] \quad (4.3.7)$$

Thus we can view I_{DS} as consisting of two components I_{DS1} and I_{DS2} .^{48,49}

$$I_{DS} = I_{DS1} + I_{DS2} \quad (4.3.8)$$

where I_{DS1} is due to the presence of drift:

$$I_{DS1} = \frac{W}{L} \int_{\psi_{s0}}^{\psi_{sL}} \mu(-Q'_t) d\psi_s \quad (4.3.9)$$

and I_{DS2} is due to the presence of diffusion:

$$I_{DS2} = \frac{W}{L} \phi_t \int_{Q'_{i0}}^{Q'_{iL}} \mu dQ'_t \quad (4.3.10)$$

The interpretation of I_{DS1} and I_{DS2} requires some caution. Note that, in general, there may not be single values of a drift current and a diffusion current in the channel, since $I_{\text{drift}}(x)$ in (4.3.3) and $I_{\text{diff}}(x)$ in (4.3.4) are functions of position.⁵⁰ Nevertheless, the above development makes it clear that I_{DS1} is there because drift was assumed to be present in the channel; were there no drift, there would be no I_{DS1} . Similarly, I_{DS2} is there because it was assumed that diffusion was present in the channel.

We now make the assumption that μ is constant along the channel (the more general case will be discussed in Sec. 4.10). Then μ can be moved outside the integral in (4.3.9) and (4.3.10), and we have

$$I_{DS1} = \frac{W}{L} \mu \int_{\psi_{s0}}^{\psi_{sL}} (-Q'_t) d\psi_s \quad (4.3.11)$$

$$I_{DS2} = \frac{W}{L} \mu \phi_t (Q'_{iL} - Q'_{i0}) \quad (4.3.12)$$

To evaluate now I_{DS1} and I_{DS2} , we need Q'_t as a function of ψ_s . An appropriate expression has been derived in (3.2.8a) and is repeated below. (The expression is, of course, assumed valid here because of the "gradual channel approximation" made in Sec. 4.1.)

$$Q'_t = -C'_{\text{ox}} \left(V_{GB} - V_{FB} - \psi_s + \frac{Q'_B}{C'_{\text{ox}}} \right) \quad (4.3.13)$$

where Q'_B is, from (3.2.5b)

$$Q'_B = -\gamma C'_{\text{ox}} \sqrt{\psi_s} \quad (4.3.14)$$

Thus Q'_I becomes

$$Q'_I = -C'_{ox}(V_{GB} - V_{FB} - \psi_s - \gamma\sqrt{\psi_s}) \quad (4.3.15)$$

Using this in (4.3.11) gives the drain current component due to the presence of drift:

$$I_{DS1} = \frac{W}{L} \mu C'_{ox} \left[(V_{GB} - V_{FB})(\psi_{sL} - \psi_{s0}) - \frac{1}{2}(\psi_{sL}^2 - \psi_{s0}^2) - \frac{2}{3} \gamma (\psi_{sL}^{3/2} - \psi_{s0}^{3/2}) \right] \quad (4.3.16)$$

Using (4.3.15) in (4.3.12) gives the drain current component due to diffusion:

$$I_{DS2} = \frac{W}{L} \mu C'_{ox} \left[\phi_t (\psi_{sL} - \psi_{s0}) + \phi_t \gamma (\psi_{sL}^{1/2} - \psi_{s0}^{1/2}) \right]^\dagger \quad (4.3.17)$$

The only step remaining is the evaluation of ψ_{s0} and ψ_{sL} from the externally applied voltages in Fig. 4.1a. Comparing this figure with Fig. 3.1c, we note that the expressions developed for the latter can be used at the source end of the channel by replacing V_{CB} by V_{SB} . Similarly, such expressions can be used at the drain end of the channel by replacing V_{CB} by V_{DB} . Thus, writing (3.2.7b) for the source end and the drain end of the channel, we obtain⁴⁰⁻⁴⁶

$$\psi_{s0} = V_{GB} - V_{FB} - \gamma\sqrt{\psi_{s0}} + \phi_t e^{(\psi_{s0} - 2\phi_F - V_{SB})/\phi_t} \quad (4.3.18a)$$

$$\psi_{sL} = V_{GB} - V_{FB} - \gamma\sqrt{\psi_{sL}} + \phi_t e^{(\psi_{sL} - 2\phi_F - V_{DB})/\phi_t} \quad (4.3.18b)$$

These equations can be solved for ψ_{s0} and ψ_{sL} by iteration. This can easily be done with a computer. Methods to speed up the computation of ψ_{s0} and ψ_{sL} can be found in the literature, where several implementations of the charge sheet model are discussed.⁵¹⁻⁵⁷

A plot of ψ_{sL} vs. V_{DB} (or of ψ_{s0} vs. V_{SB}) is shown in Fig. 4.5. The symbols used on the figure are defined in Chap. 3. The plot saturates at a value dependent on V_{GB} , as given by (3.2.16). The numbered points are in relation to an upcoming discussion.

Note that by neglecting the edge effects near the drain we have ended up with a potential ψ_{sL} at the drain end of the channel which will, in general, be different from the potential in the n^+ region. Thus it would seem that a potential discontinuity would exist at the n^+ region boundary. As expected from the associated discussion in Sec. 3.5, a more detailed picture would include a transition region containing the boundary, over which the potential would change continuously and, eventually, would become equal to

[†]The reader is warned that this equation, although theoretically sound, can cause numerical difficulties in weak inversion. This is explained later in this section. An improvement is suggested in Prob. 4.2.

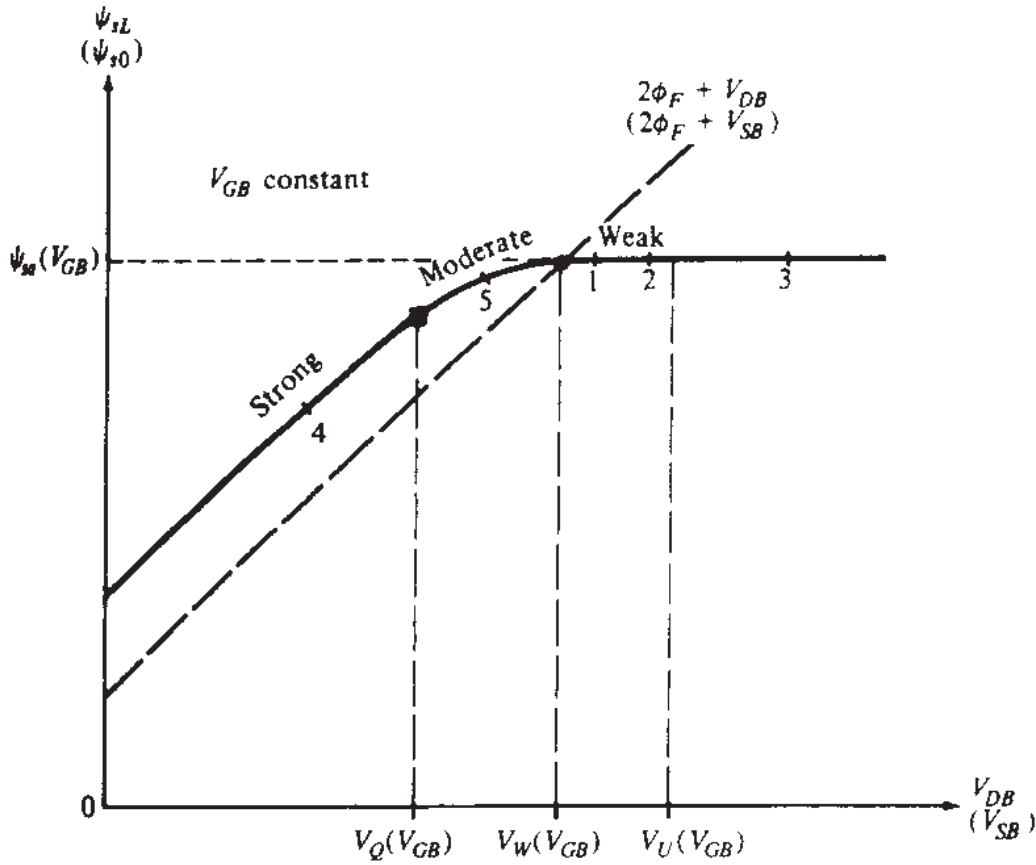


FIGURE 4.5

Surface potential at the drain (source) end of the channel vs. drain-substrate (source-substrate) voltage for a given gate-substrate voltage.

the potential in the n^+ region. The length of this transition region may be expected to be roughly the same as the depth of the depletion region under the drain. In this region, one cannot assume that the electric field is practically vertical. Similar comments apply to the source end of the channel. Based on such a picture, one could interpret L in the above analysis as representing the length of the channel, excluding the source and the drain transition regions, and ψ_{s0} and ψ_{sL} as the surface potentials to the right and left of these regions, respectively. Then (4.3.18) would again be used, assuming the effect of V_{SB} and V_{DB} at the two points remained essentially the same.[†] However, for the long channels assumed in this chapter we can continue viewing L as the total source-drain spacing. Note that in our analysis we are forced to neglect the above effects since they cannot be handled by the “gradual channel approximation,” which we have adopted to be able to obtain simple analytical results. An investigation of the above effects would have made necessary the use of two-dimensional analysis requiring a numerical solution, or at least a pseudo-two-dimensional analysis (Chap. 6).

We now fix V_{SB} , keep V_{GB} as a parameter, and vary V_{DB} . We find the corresponding ψ_{s0} and ψ_{sL} from (4.3.18) and substitute them in (4.3.16) and (4.3.17).

[†]Readers familiar with quasi-Fermi levels (see Appendixes A and H) will recognize this assumption as equivalent to assuming that the electron quasi-Fermi levels remain essentially constant over the transition regions; a corresponding assumption is often made in pn junction theory.^{16,58}

Adding I_{DS1} and I_{DS2} produces I_{DS} as in (4.3.8), and gives the plots of Fig. 4.2a, where the current is shown vs. V_{DB} , with V_{GB} as a parameter. These plots can be converted to the ones in Fig. 4.2b, where we show the current vs. $V_{DS} = V_{DB} - V_{SB}$, with $V_{GS} = V_{GB} - V_{SB}$ as a parameter. Plotting $\ln I_{DS}$, rather than I_{DS} , produces Fig. 4.3. In these figures, the regions of operation are marked according to Table 4.1, with the values of V_{GB} or V_{GS} delineating these regions denoted by the symbols defined in Chap. 3. These values are not of interest at this point. What is important is to note that the *single* expression (4.3.8) predicts the current in all these regions. Experiments agree well with this expression.^{43-46,51-56}

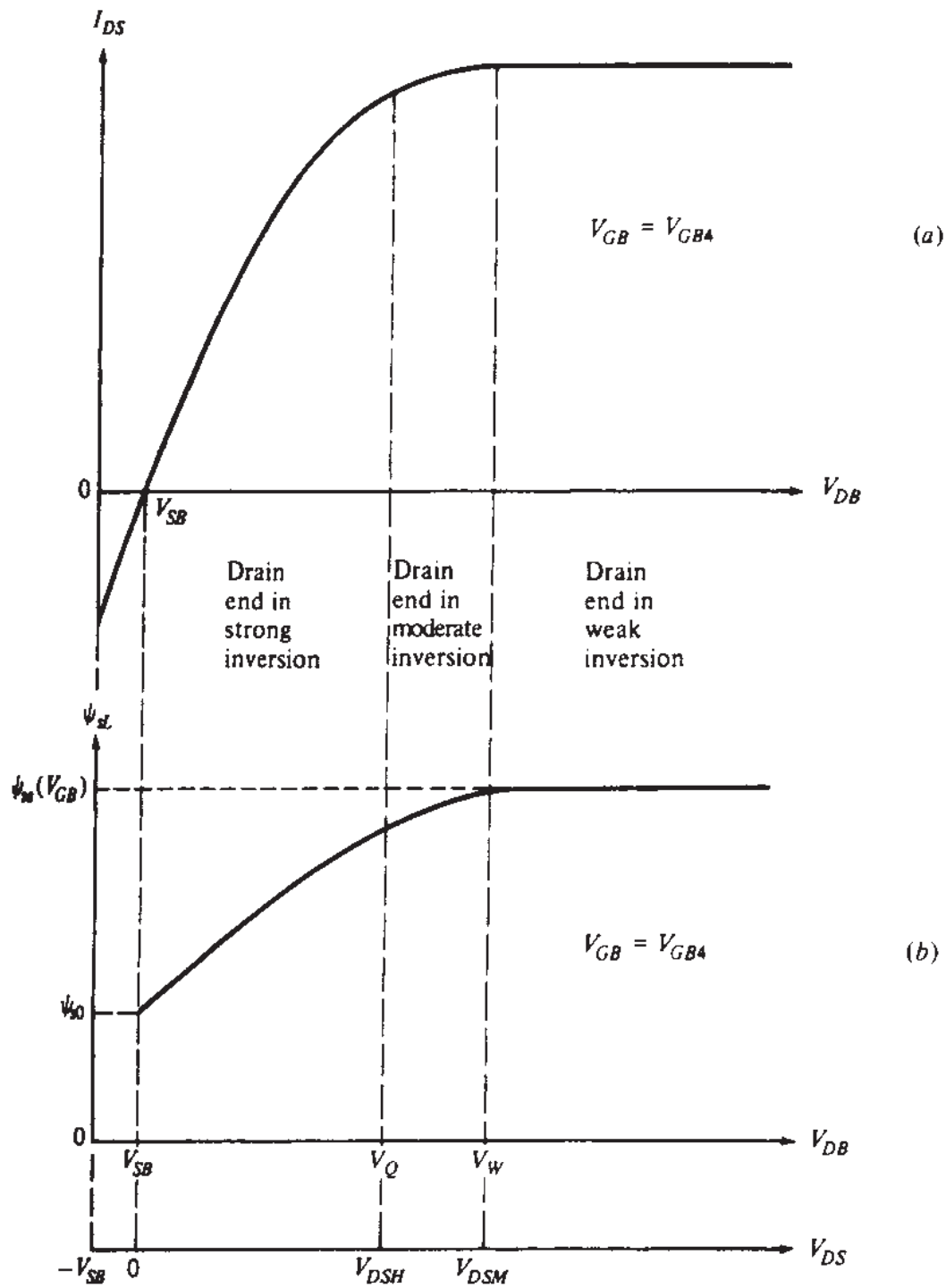
Note that all curves in Fig. 4.2a saturate for large V_{DB} . This can be understood by relating the drain current for one of the V_{GB} values (say, V_{GB4} in Fig. 4.2a) to the corresponding surface potential ψ_{sL} at the drain. This is shown in Fig. 4.6: increasing V_{DB} eventually drives the drain end of the channel into weak inversion, where ψ_{sL} becomes practically constant at a value dependent on $V_{GB} = V_{GB4}$ only, for reasons explained in Chap. 3 [see (3.2.16) and Fig. 3.12]. Increasing V_{DB} further has little effect on ψ_{sL} . Thus I_{DS1} and I_{DS2} in (4.3.16) and (4.3.17) also become independent of V_{DB} . It is seen then that, although the source end of the channel can be, say, strongly inverted with a large $|Q'_I|$ and with ψ_{s0} strongly dependent on V_{SB} , the drain end of the channel can be weakly inverted with a small $|Q'_I|$ and with ψ_{sL} practically independent of V_{DB} . However, the current at the two ends (and throughout the channel) is the same. This should not be surprising since, as follows from (4.3.5), the current at any point in the channel is not determined only by the value of Q'_I at that point. This will be discussed further later on.

Consider now a value of V_{DB} corresponding to the saturation part of the curves in Fig. 4.2 or 4.3. If I_{DS} and its components are plotted for that value vs. V_{GB} , we obtain the curves of Fig. 4.7.⁴⁹ To include a large range of currents, a logarithmic vertical axis is used. The regions of inversion across the horizontal axis have been marked according to the definitions of Table 4.1. Here V_{DB} is larger than V_{SB} , so that the more heavily inverted channel end is the one next to the source. The region of inversion at that end can be determined as in Chap. 3. We note here that in transistor literature the weak inversion region is often not bounded from below; i.e., everything below moderate inversion is called weak inversion, as long as the current does not become so low as to be masked by leakage currents (Sec. 4.6).

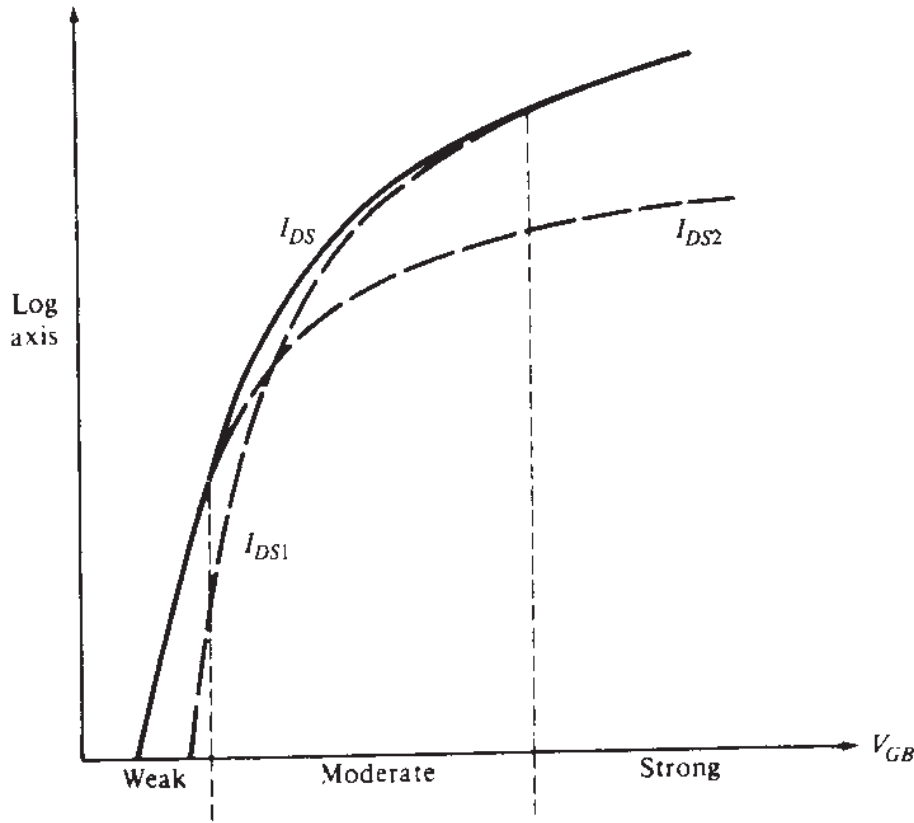
It is seen in Fig. 4.7 that in strong inversion $I_{DS} \approx I_{DS1}$, so the current is mainly due to the presence of drift. In weak inversion, the current is mainly due to the presence of diffusion, since $I_{DS} \approx I_{DS2}$. However, in moderate inversion *both* I_{DS1} and I_{DS2} are important; both drift and diffusion play an important role in this region. Similar conclusions can be reached for other V_{DB} values.

SYMMETRY. It is clear from (4.3.8), (4.3.16), and (4.3.17) that I_{DS} can be written in the form

$$I_{DS} = \frac{W}{L} [f(\psi_{sL}) - f(\psi_{s0})] \quad (4.3.19)$$

**FIGURE 4.6**

(a) Drain-source current for a transistor operating in strong inversion vs. drain-substrate voltage for fixed source-substrate and gate-substrate voltages; (b) surface potential at the drain end of the channel, corresponding to (a). A drain-source voltage axis is shown at the bottom of the figure.

**FIGURE 4.7**

Drain-source current I_{DS} , its component due to drift I_{DS1} , and its component due to diffusion I_{DS2} vs. gate-substrate voltage for a large drain-substrate voltage corresponding to the saturation part of the characteristics in Fig. 4.2a.⁴⁹ A logarithmic axis is used for the current.

where

$$f(\psi_s) = \mu C'_{ox} \left[(V_{GB} - V_{FB} + \phi_t) \psi_s - \frac{1}{2} \psi_s^2 - \frac{2}{3} \gamma \psi_s^{3/2} + \phi_t \gamma \psi_s^{1/2} \right] \quad (4.3.20)$$

Equation (4.3.19) is in a form which emphasizes the symmetry of the transistor. If the potentials at the source and drain are interchanged, the only difference will be that I_{DS} will change sign. That I_{DS} must be in the form of (4.3.19) can also be deduced directly from (4.3.7). In fact, this form, with $f(\psi_s)$ an appropriate function, will be valid *even* if μ depends on ψ_s in that equation.

NUMERICAL ISSUES. We now come to a numerical problem associated with (4.3.17) in weak inversion, where $I_{DS} \approx I_{DS2}$. Consider a transistor operating in this region, with V_{SB} and V_{DB} corresponding to two points between V_w and V_U in Fig. 4.5. As seen, ψ_{sL} is then nearly equal to ψ_{s0} (no difference can be seen in the figure). Thus, even a very small error in the values of ψ_{sL} and ψ_{s0} can mean a large relative error in the differences on which (4.3.17) relies; thus, the resulting relative error in the current can also be large. One can then hope to get I_{DS} from (4.3.17) only if ψ_{sL} and ψ_{s0} are known extremely accurately, which requires several iterations when solving

(4.3.18). Approximate explicit expressions for ψ_s will *not* work with (4.3.17) in weak inversion. (A way to circumvent this drawback is suggested in Prob. 4.2.) We also note that high-precision arithmetic is required in evaluating I_{DS2} from (4.3.17).†

The general charge sheet model we have presented is often developed in the literature by using the concept of quasi-Fermi potentials (Appendix J). This results in a more complex derivation^{40,41,43-46} than the one we have presented. It is also possible to calculate the current without making the charge sheet approximation, allowing for the spreading of the inversion layer below the surface, and even allowing for the presence of holes in the depletion region; this was done quite early by Pao and Sah¹⁰ (Appendix J). Although their analysis is recognized to be very general and accurate, it includes the numerical evaluation of a double integral and is, thus, computationally inefficient. It has been shown that the double-integral formulation can be reduced to an equivalent single-integral one,⁴⁷ but numerical integration is still required. The charge sheet model we have presented is known to be in excellent agreement with these general formulations and does not require numerical integration.

SURFACE POTENTIAL AND INVERSION LAYER CHARGE VS. POSITION. For visualizing transistor operation and for calculating certain quantities, it will be useful to relate the surface potential ψ_s to the position along the channel x . The part of the device in Fig. 4.1a between point x and the source can be viewed as a transistor by itself, with point x playing the role of the “drain” and $\psi_s(x)$ playing the role of the surface potential at the “drain” end; the channel length of this transistor will be x . Thus, in lieu of (4.3.19), we will have for this transistor

$$I_{DS} = \frac{W}{x} [f(\psi_s(x)) - f(\psi_{s0})] \quad (4.3.21)$$

where, of course, the current is the same as in the complete device owing to current continuity in the channel. Eliminating I_{DS} between the above equation and (4.3.19), we obtain

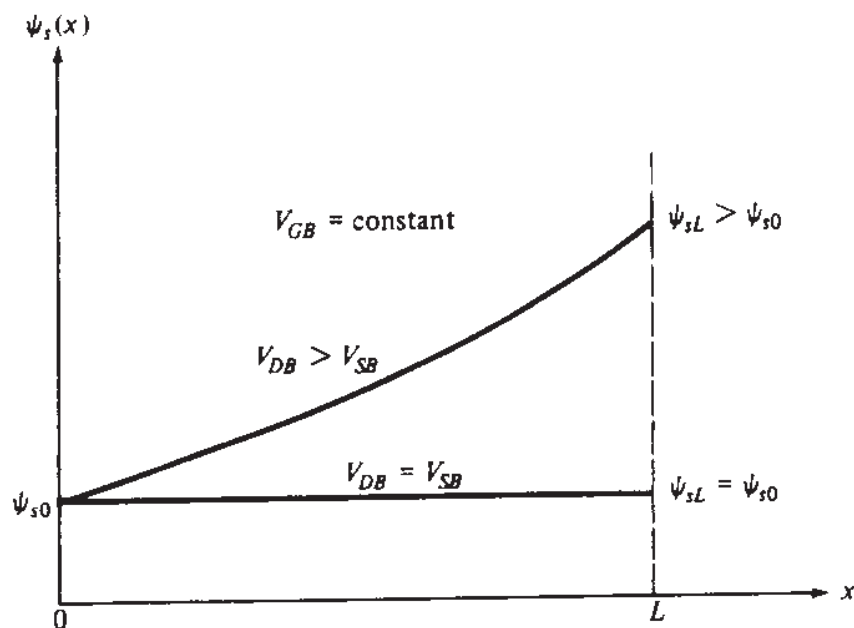
$$\frac{x}{L} = \frac{f(\psi_s(x)) - f(\psi_{s0})}{f(\psi_{sL}) - f(\psi_{s0})} \quad (4.3.22)$$

This equation gives the relation between x and $\psi_s(x)$. The easy way to get results from it is to give values to $\psi_s(x)$ between ψ_{s0} and ψ_{sL} , and determine x . In strong inversion, plots obtained in this way have the form shown in Fig. 4.8. In moderate

†A numerically robust form for $I_{DS} = I_{DS1} + I_{DS2}$ can be obtained⁵⁹ by factoring out $\psi_{sL} - \psi_{s0}$ from (4.3.16) and (4.3.17). This gives

$$I_{DS} = \frac{W}{L} \mu C_{ox} \left[V_{GB} - V_{FB} - \frac{1}{2}(\psi_{sL} + \psi_{s0}) - \frac{2}{3} \gamma \frac{\psi_{sL} + \psi_{sL}^{1/2} \psi_{s0}^{1/2} + \psi_{s0}}{\psi_{sL}^{1/2} + \psi_{s0}^{1/2}} + \phi_t \left(1 + \frac{\gamma}{\psi_{sL}^{1/2} + \psi_{s0}^{1/2}} \right) \right] (\psi_{sL} - \psi_{s0})$$

This form is less sensitive than (4.3.16) and (4.3.17) to limitations caused by finite precision arithmetic.

**FIGURE 4.8**

Surface potential vs. distance from the source for given V_{GB} and V_{SB} , for two values of V_{DB} (strong inversion is assumed).

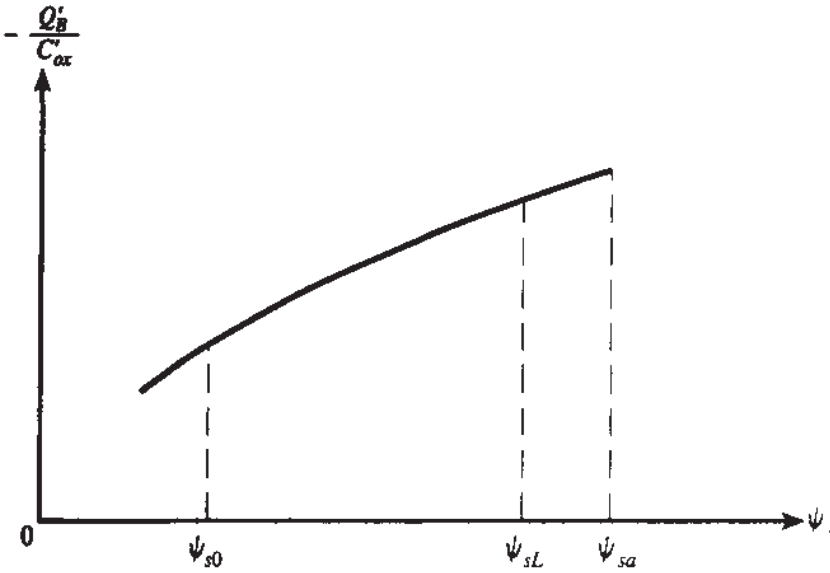
inversion, the variation of ψ_s with x for $V_{DB} > V_{SB}$ would be less pronounced. In weak inversion, the two curves would practically coincide since, in that region, the variations of the surface potential along the channel is negligible even when $V_{DB} > V_{SB}$.

The variation of charges with position in the channel can now be determined. One can, for example, consider (4.3.15) and (4.3.22) as a parametric representation of Q'_I vs. x . For each value of $\psi_s(x)$ (between ψ_{s0} and ψ_{sL}) given to (4.3.22) and (4.3.15), a point (x, Q'_I) is obtained. Thus one can plot Q'_I vs. x . For the case illustrated by the upper curve of Fig. 4.8, one finds from (4.3.15) that $|Q'_I|$ as predicted by the charge sheet model decreases monotonically along the channel as we go from the source toward the drain.

One can similarly evaluate $d\psi_s/dx$ and dQ'_I/dx as a function of position x , and substitute them into (4.3.3) and (4.3.4). It is found that $I_{\text{drift}}(x)$ decreases with x , whereas $I_{\text{diff}}(x)$ increases.⁵⁰ The values of these two current components at every x are such that their sum is constant and equal to I_{DS} .

4.3.2 Simplified Charge Sheet Models

The "complete" charge sheet model of (4.3.8) and (4.3.16) to (4.3.18) is very accurate, but is too complicated for some applications; this becomes apparent when the same formulation is extended to charge evaluation, which is needed in transient response calculations (Chap. 7). Part of this complexity is due to the $3/2$ power and $1/2$ power terms in (4.3.16) and (4.3.17), respectively. It is clear from the development presented that the origin of these terms is the square root in (4.3.15). This square root is, in turn, due to the expression for Q'_B in (4.3.14). Figure 4.9 shows a plot of

**FIGURE 4.9**

The quantity $-Q'_B/C'_{ox}$ vs. surface potential.

$-Q'_B/C'_{ox}$ vs. ψ_s from that equation. The surface potential in the channel will be between ψ_{s0} and ψ_{sL} (see our discussion above on surface potential vs. position, and Fig. 4.8). As seen in Fig. 4.5, the surface potential at the drain, ψ_{sL} , can be no higher than ψ_{sa} , and in fact it can only approach the latter asymptotically, for very large drain-substrate voltages. Since the slope of the plot in Fig. 4.9 does not vary much, it is reasonable to attempt to approximate $-Q'_B/C'_{ox}$ by the first two terms of its Taylor expansion around a convenient expansion point $\psi_s = \psi_{se}$ (choices for ψ_{se} will be discussed shortly). This gives, from (4.3.14):

$$-\frac{Q'_B}{C'_{ox}} = \gamma\sqrt{\psi_{se}} + \frac{\gamma}{2\sqrt{\psi_{se}}}(\psi_s - \psi_{se}) \quad (4.3.23)$$

Let us define, for later use, a quantity α as follows:

$$\alpha = 1 + \frac{\gamma}{2\sqrt{\psi_{se}}} \quad (4.3.24)$$

In terms of this quantity, (4.3.23) can be written as

$$-\frac{Q'_B}{C'_{ox}} = \gamma\sqrt{\psi_{se}} + (\alpha - 1)(\psi_s - \psi_{se}) \quad (4.3.25)$$

Using this in (4.3.13), gives the inversion layer charge:

$$Q'_I = -C'_{ox}[V_{GB} - V_{FB} - \psi_{se} - \gamma\sqrt{\psi_{se}} - \alpha(\psi_s - \psi_{se})] \quad (4.3.26)$$

We see that under the above approximation, Q'_I for a fixed V_{GB} varies linearly⁶⁰ with variations in ψ_s . This approximation for Q'_I is more satisfactory than it is for Q'_B , as Q'_I already contained a linear term to begin with [see (4.3.13)]. The nearly linear

variation of Q'_I with ψ_s for a fixed V_{GB} can be also be observed in Fig. 3.11. Differentiation of (4.3.26) gives

$$\frac{dQ'_I}{d\psi_s} = \alpha C'_{ox} \quad (4.3.27)$$

This is a key relation for simplifying the expression for the current.⁶⁰ Indeed, using it in (4.3.11) to perform a change of variables from ψ_s to Q'_I , we obtain

$$I_{DS1} = \frac{W}{L} \mu \int_{Q'_{I0}}^{Q'_{IL}} (-Q'_I) \frac{1}{\alpha C'_{ox}} dQ'_I \quad (4.3.28)$$

which gives:

$$I_{DS1} = \frac{W}{L} \frac{\mu}{2\alpha C'_{ox}} (Q'^2_{I0} - Q'^2_{IL}) \quad (4.3.29)$$

whereas I_{DS2} is still given by (4.3.12), repeated below:

$$I_{DS2} = \frac{W}{L} \mu \phi_1 (Q'_{IL} - Q'_{I0}) \quad (4.3.30)$$

Adding these two expressions gives a result similar to ones derived by various means elsewhere.^{60,70,71,64,65,62} Ways to evaluate Q'_{IL} and Q'_{I0} , either iteratively or using explicit approximate expressions in terms of external bias voltages, have been proposed.^{70,71,64,65}

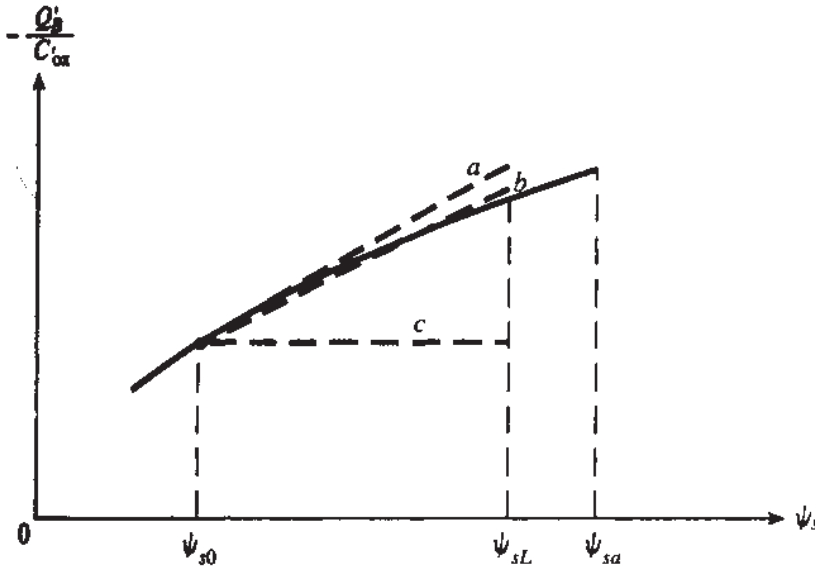
In the above equations, we can use Q'_{I0} and Q'_{IL} as obtained from (4.3.26), with $\psi_s = \psi_{s0}$ and $\psi_s = \psi_{sL}$, respectively. This gives

$$I_{DS1} = \frac{W}{L} \mu C'_{ox} \left[(V_{GB} - V_{FB} - \gamma \sqrt{\psi_{se}} - \psi_{se} + \alpha \psi_{se}) (\psi_{sL} - \psi_{s0}) - \frac{\alpha}{2} (\psi_{sL}^2 - \psi_{s0}^2) \right] \quad (4.3.31)$$

$$I_{DS2} = \frac{W}{L} \mu C'_{ox} \alpha \phi_1 (\psi_{sL} - \psi_{s0}) \quad (4.3.32)$$

Although it is tempting to further simplify (4.3.31) by taking advantage of the form of (4.3.24), we will not do so as the value of α will have to be modified in some cases, as we will see.

†For improved numerical robustness $(\psi_{sL} - \psi_{s0})$ can be factored out.

**FIGURE 4.10**

Solid line: $-Q'_D/C'_{ox}$ vs. surface potential. Line *a*: approximation (4.3.25) with (4.3.33) and (4.3.36). Line *b*: improved approximation obtained by slightly lowering the value of α . Line *c*: approximation $\alpha = 1$.

Our discussion so far has been in terms of an arbitrary expansion point, $\psi_s = \psi_{se}$. Two choices for ψ_{se} have been proposed in the literature. These are considered below.

1. EXPANSION AROUND ψ_{s0} : SOURCE-REFERENCED MODELS. Let us choose:^{61,62}

$$\psi_{se} = \psi_{s0} \quad (4.3.33)$$

This corresponds to approximating $-Q'_D/C'_{ox}$ [see (4.3.23)–(4.3.25)] as shown by line *a* in Fig. 4.10. This choice of ψ_{se} was originally suggested for a strong inversion model in Ref. 24. Using the above value in (4.3.31) and (4.3.32) we obtain

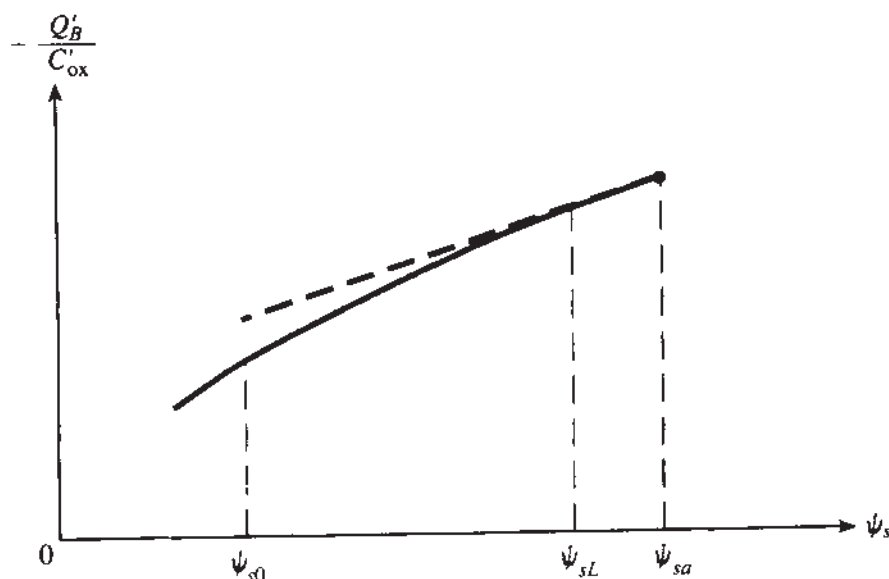
$$I_{DS1} = \frac{W}{L} \mu C'_{ox} \left[(V_{GB} - V_{FB} - \psi_{s0} - \gamma \sqrt{\psi_{s0}})(\psi_{sL} - \psi_{s0}) - \frac{\alpha}{2} (\psi_{sL} - \psi_{s0})^2 \right] \quad (4.3.34)$$

$$I_{DS2} = \frac{W}{L} \mu C'_{ox} \phi_t \alpha (\psi_{sL} - \psi_{s0}) \quad (4.3.35)$$

where the nominal value to be used for α , which we will denote by α_1 , can be found by using (4.3.33) in (4.3.24):

$$\alpha = \alpha_1 = 1 + \frac{\gamma}{2\sqrt{\psi_{s0}}} \quad (4.3.36)$$

As seen, in the current equations the drain surface potential appears only in a *difference* with the source surface potential. This is typical of *source-referenced* models

**FIGURE 4.11**

Solid line: $-Q'_B/C'_{ox}$ vs. surface potential. Broken line: approximation (4.3.23) with (4.3.37).

and will be seen to lead to popular strong-inversion models in Sec. 4.5. On the other hand, the inherent symmetry of the ideal MOS transistor is not apparent in the above equations. Certain refinements are needed before the above approach can lead into a fully developed model.[†]

It has been seen that (4.3.25), with α as in (4.3.36), results in the approximation shown by the upper broken line in Fig. 4.10. One can improve the overall accuracy of Q'_B vs. ψ_s , by replacing α above by a slightly smaller value. This corresponds to the type of approximation shown by line *b* in Fig. 4.10. We will see examples of this practice in Sec. 4.5. The slope of line *b* can be made a function of $\psi_{sL} - \psi_{s0}$ for even better accuracy.⁶³ For future use, Fig. 4.10 also shows the poor approximation resulting from assuming $\alpha = 1$ in (4.3.25) (line *c*).

2. EXPANSION AROUND ψ_{sa} : SYMMETRIC MODELS. A different choice for ψ_{se} is^{64,65}

$$\psi_{se} = \psi_{sa} \quad (4.3.37)$$

This leads to the approximation shown by the broken line in Fig. 4.11. This choice gives an accurate Q'_B , and thus an accurate Q'_I , when $\psi_s \approx \psi_{sa}$. It will be recalled that at such values of ψ_s , Q'_I becomes negligible in comparison to Q'_B (Fig.

[†]As seen from line *a* in Fig. 4.10, the choice in (4.3.33) results in an expansion which is most accurate near the source. Near the drain, Q'_B and thus Q'_I will be somewhat in error. In the complete charge sheet model, as V_{DB} was raised, ψ_{sL} approached ψ_{sa} and Q'_{IL} approached 0, leading to "saturation" of the I_{DS} curves as in Fig. 4.6a. Here, though, because of the above small error Q'_{IL} approaches 0 at a value of ψ_{sL} slightly different from ψ_{sa} . This should be taken into account.⁶¹ It turns out that this is not a serious problem, as for real devices the saturation value of ψ_{sL} has to be modified anyway⁶² to take into account other effects such as velocity saturation (Chap. 6).

3.11), and we are in weak inversion or depletion. Thus, this approximation provides best accuracy for Q'_B when Q'_B is dominant. At the other extreme, i.e., when Q'_I is dominant, inaccuracies in Q'_B matter little anyway, since this charge contributes a negligible amount to the total semiconductor charge.⁶⁵ Using (4.3.37) in (4.3.24), one finds that a α becomes equal to the quantity n in (3.2.18):

$$\alpha = n = 1 + \frac{\gamma}{2\sqrt{\psi_{sa}}} \quad (4.3.38)$$

Using (4.3.37) and (4.3.38) in (4.3.31) and (4.3.32), we obtain

$$I_{DS1} = \frac{W}{L} \mu C'_{ox} \left[\left(V_{GB} - V_{FB} - \frac{\gamma}{2} \sqrt{\psi_{sa}} \right) (\psi_{sL} - \psi_{s0}) - \frac{n}{2} (\psi_{sL}^2 - \psi_{s0}^2) \right] \quad (4.3.39)$$

$$I_{DS2} = \frac{W}{L} \mu C'_{ox} n \phi_i (\psi_{sL} - \psi_{s0}) \quad (4.3.40)$$

These equations make apparent the symmetric role of the source and drain, just like the complete charge sheet model in Sec. 4.3.1. The values of ψ_{s0} and ψ_{sL} can be calculated, for given V_{GB} , V_{SB} , and V_{DB} , from (4.3.18).

FORWARD AND REVERSE CURRENTS. From (4.3.8), (4.3.29), and (4.3.30) we can write [below we assume that $\alpha = n$, from (4.3.38)]:

$$I_{DS} = I_{DS1} + I_{DS2} \quad (4.3.41a)$$

$$= \frac{W}{L} \mu \left[\frac{1}{2nC'_{ox}} (Q'^2_{I0} - Q'^2_{IL}) + \phi_i (Q'_{IL} - Q'_{I0}) \right] \quad (4.3.41b)$$

$$= \frac{W}{L} \mu \left(\frac{Q'^2_{I0}}{2nC'_{ox}} - \phi_i Q'_{I0} \right) - \frac{W}{L} \mu \left(\frac{Q'^2_{IL}}{2nC'_{ox}} - \phi_i Q'_{IL} \right) \quad (4.3.41c)$$

This equation can be written in a form used in Refs. 66 to 69:

$$I_{DS} = I_F - I_R \quad (4.3.42)$$

where:^{64,65}

$$I_F = \frac{W}{L} \mu \left(\frac{Q'^2_{I0}}{2nC'_{ox}} - \phi_i Q'_{I0} \right) \quad (4.3.43)$$

$$I_R = \frac{W}{L} \mu \left(\frac{Q'_{IL}{}^2}{2nC'_{ox}} - \phi_i Q'_{IL} \right) \quad (4.3.44)$$

In the saturation region, as V_{DB} becomes large and ψ_{sL} approaches ψ_{sa} , Q'_{IL} approaches 0 and so does I_R (see Fig. 3.12). The drain current then approaches a constant value, as expected. We thus have

$$I_{DS, \text{saturation}} = I_F \quad (4.3.45)$$

Similarly, if V_{SB} becomes large and ψ_{s0} approaches ψ_{sa} , Q'_{I0} approaches 0 and so does I_F . We can interpret this as *reverse saturation*. We thus have:

$$I_{DS, \text{rev. saturation}} = -I_R \quad (4.3.46)$$

We see then that the drain current in (4.3.42) can be interpreted^{66-69,64} as the sum of a *forward* component, I_F , and a *reverse* component, I_R . The first of these components depends only on V_{GB} and V_{SB} ; the second depends only on V_{GB} and V_{DB} . A similar interpretation is possible for the complete charge-sheet model of Sec. 4.3.1.

CURRENT-BASED MODEL. It was seen above that I_{DS} can be expressed in terms of the forward saturation current I_F and the reverse saturation current I_R . Similarly, other quantities of interest in the transistor, such as charges and small-signal parameters, can also be expressed in terms of these currents.^{66-69,72} For example, one can express⁷² Q'_{I0} and Q'_{IL} in terms of I_F and I_R , by solving (4.3.43) and (4.3.44). One thus ends up with a model in which all quantities of interest, instead of being functions of bias voltages, are functions of I_F and I_R . Since these are, respectively, the forward and reverse saturation currents, they are easily measurable and are the currents that are often used to bias the transistors in practical circuit design. The reader is referred to the literature for further information on this concept.^{66-69,72}

4.3.3 Model Based on Quasi-Fermi Potentials

All of the above models are based on (4.3.5), which expresses the total channel current as a sum of a drift and a diffusion term, proportional to $Q'_i(d\psi_s/dx)$ and dQ'_i/dx , respectively. It can be shown that these two terms can be combined into a single term proportional to $Q'_i(dV/dx)$, where V is the so-called "quasi-Fermi potential." This formulation is outlined in Appendix J and is quite general, in that it is also valid without making the approximation that the inversion layer is a charge sheet of negligible thickness.¹⁰ Unfortunately, the models this formulation has lead to are complicated (Appendix J).

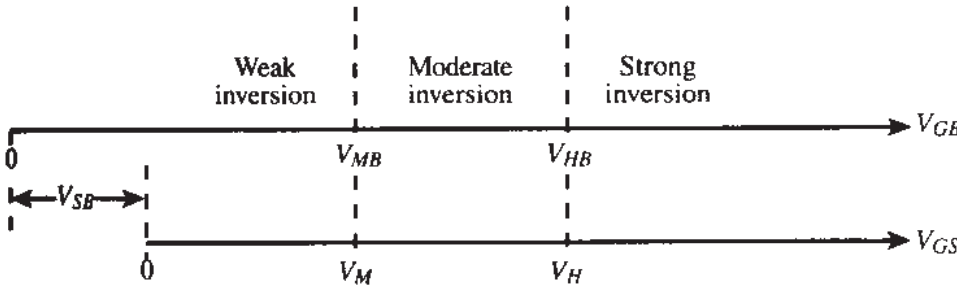


FIGURE 4.12

Definitions of regions of inversion in terms of gate voltage, for a given V_{SB} (it is assumed that $V_{DB} \geq V_{SB}$).

4.4 REGIONS OF INVERSION IN TERMS OF TERMINAL VOLTAGES

It was seen in Sec. 4.2 that the transistor is said to operate in a certain region of inversion, if its most heavily inverted channel end is in that region of inversion. On the basis of this definition, the regions of inversion have been defined as in Table 4.1. We now define these regions in terms of externally applied voltages.

DEFINITIONS IN TERMS OF GATE VOLTAGE. Let us assume arbitrarily that the drain end is no more heavily inverted than the channel end. This is equivalent to assuming:†

$$V_{DB} \geq V_{SB} \quad (4.4.1)$$

This assumption can be written in terms of V_{DS} as follows:

$$V_{DS} \geq 0 \quad (4.4.2)$$

Then we can give the condition for operation in a certain inversion region, in terms of the inversion level at the *source* end of the channel. We can use appropriate conditions developed in Chap. 3 for the structure of Fig. 3.1c or d, simply by replacing terminal C by the source terminal S. Thus, from Fig. 3.2 or Table 3.2 (second and third rows), the regions can be defined in terms of V_{GB} or V_{GS} as shown in Fig. 4.12 (in drawing this figure, we have arbitrarily assumed that V_M is greater than 0). We remind the reader that the quantities marked on the V_{GS} axis are given by (Table 3.1):

$$V_M = V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F + V_{SB}} \quad (4.4.3)$$

$$V_H = V_M + V_Z \quad (4.4.4)$$

†We make this convenient assumption just in order to be able to draw some figures and give some equations in a specific way. This does not really restrict the generality of our discussion. If $V_{DB} < V_{SB}$ ($V_{DS} < 0$), the role of the source and drain are interchanged; then simply replace V_{DB} by V_{SB} and vice versa, and V_{DS} by V_{SD} , in the equations that follow. See Prob. 4.7 for other ways to define regions of operation.

with the value of V_Z being several tenths of a volt (0.5 to 0.6 V at room temperature for practical devices; V_Z decreases slightly with increasing V_{SB}). The corresponding V_{GB} limits can be found by simply adding V_{SB} to the above values, as indicated in Fig. 4.12.

Recall that in Chaps. 2 and 3, a *lower* limit for the weak inversion region was also defined at $V_{GB} = V_{LB}$. However, when dealing with a complete transistor, this lower limit is rather academic, for two reasons: (1) Nothing really special happens at that limit. As seen in Fig. 3.2, the various plots are smooth, and their shape does not change drastically below $V_{GB} = V_{LB}$. For this reason, some of the equations we will develop in this section can hold even below that limit. (2) In a real MOS transistor, the observable (and usable) drain current is the sum of the real drain current and a reverse-junction leakage current (which can include the leakage of the drain-substrate n^+p junction as well as leakage across the depletion region under the channel). Thus, a "pragmatic" lower limit for weak-inversion operation can be taken to be at a point where the leakage current can be neglected, requiring

$$I_D \gg I_j \quad (4.4.5)$$

where I_j is the leakage current. This current is difficult to predict; it depends on fabrication details and increases rapidly with increasing temperature. Often, at room temperatures and above, (4.4.5) will restrict the usable weak-inversion region width to a value smaller than $V_{MB} - V_{LB}$.

DEFINITIONS IN TERMS OF SOURCE AND DRAIN VOLTAGES. For a given V_{GB} , the surface potential at the drain (or source) end of the channel can be plotted vs. the drain (or source) voltage with respect to the substrate, as shown in Fig. 4.5. It is clear that one can determine the level of inversion at the source and drain ends of the channel by comparing each of V_{SB} and V_{DB} to the quantities V_Q and V_W , which were discussed in Sec. 3.5 and given in the last row of Table 3.1. For convenience, we repeat the values of these quantities below:

$$V_Q = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB} - V_Z} \right)^2 - 2\phi_F \quad (4.4.6)$$

$$V_W = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right)^2 - 2\phi_F \quad (4.4.7)$$

where, in (4.4.6), V_Z is the same quantity discussed above.

Thus, for example, it follows from Table 4.1 that if the conditions at the source and drain correspond respectively to points 1 and 2 in Fig. 4.5, the transistor operates in weak inversion, and the same will be true for points 1 and 3. For points, respectively, 5 and 2 or 5 and 3, it operates in moderate inversion. The transistor operates in strong inversion for points 4 and 5, or 4 and 1, or 4 and 3, etc. Note that if point 3 corresponds to the source and point 4 to the drain, the transistor is *still* said to operate

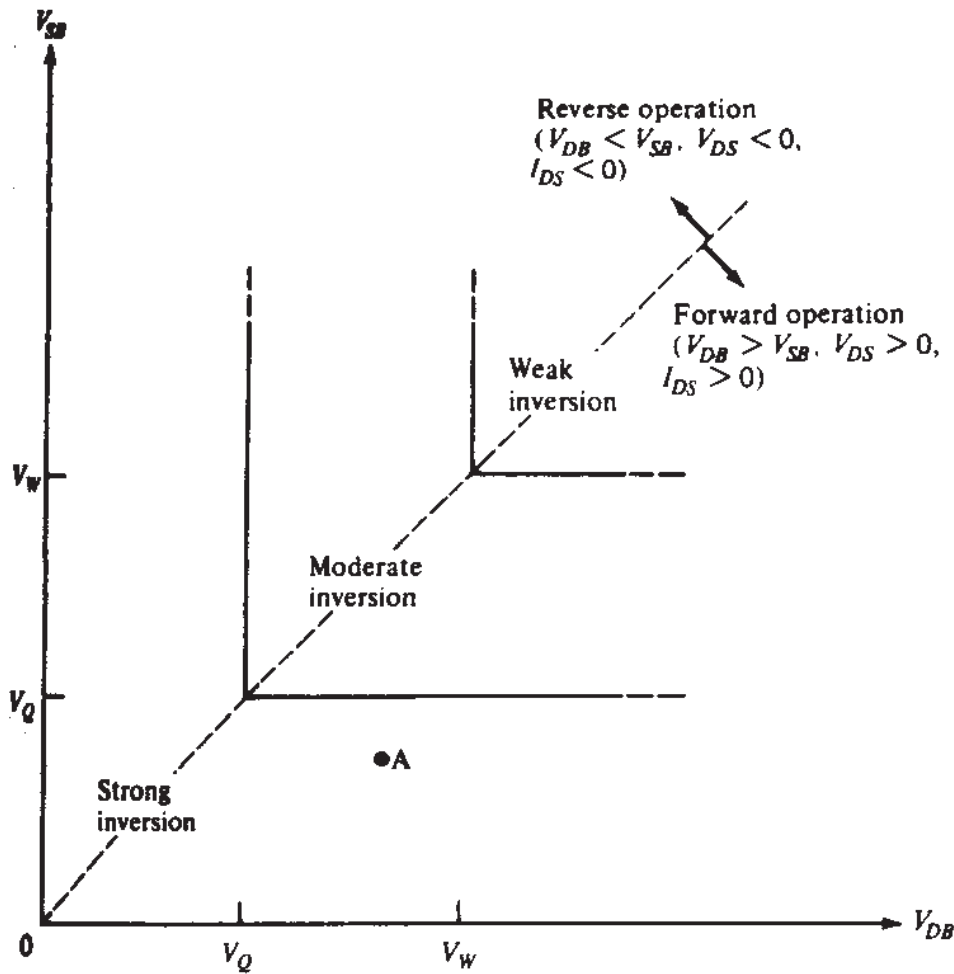


FIGURE 4.13

Definition of regions of operation for a MOS transistor in terms of source-substrate and drain-substrate voltages for a given V_{GB} .

in strong inversion, according to Table 4.1. However, now the most heavily inverted channel end is the one next to the drain; here $V_{DB} < V_{SB}$ ($V_{DS} < 0$), which results in $I_{DS} < 0$. This sort of operation can be termed as *reverse*, whereas *forward* operation can refer to the opposite case, for which most of our equations have been written.

Using the above observations, it is easy to see that a general way to define the regions of inversion is as shown in Fig. 4.13. As an example, a transistor with $V_{SB} < V_Q$ and $V_Q < V_{DB} < V_W$ can correspond to, say, point A. The transistor is then in (forward) strong inversion.

Note that in accordance with the discussion above, we prefer not to specify a limit between weak inversion and depletion. Rather, we take a transistor to operate in weak inversion if it corresponds to a point in the upper right-hand corner in Fig. 4.13 as long as the current remains well above the leakage current.

In the next three sections, we will discuss each region of inversion separately, looking for possible simplified expressions for the drain-source current. Such simplification is sought for several reasons. An obvious one is computational speed. Having to solve (4.3.18) numerically makes the general charge sheet model unsuitable for

hand calculations and not very efficient for computer simulation of large circuits. Another reason for seeking simplification is that, by concentrating on the dominant phenomena in each region of inversion and by making certain approximations, the resulting form of the drain current expression can “display” these phenomena in an obvious manner in some cases. Certain easily identifiable or measurable parameters can be made to appear in the expression. Finally, if the form of the functional dependence of the drain current on the terminal voltages is made apparent in a simple, explicit expression, a circuit designer may be able to take advantage of this form in creating new circuits.

4.5 STRONG INVERSION

A transistor is said to operate in strong inversion if at least one of the two channel ends is strongly inverted. Several models have been developed for the strong inversion region. Each has something going for it. Historically, each strong inversion model was developed from ground zero, starting with appropriate strong inversion assumptions. However, in order to make clear where these models fit in the overall modeling framework, we will develop them starting from the charge sheet models of Sec. 4.3.

4.5.1 Complete Symmetric Strong-Inversion Model

NONSATURATION. With strong inversion guaranteed at the source end of the channel, if $V_{DB} = V_{SB}$ ($V_{DS} = 0$), the drain end will also be strongly inverted. If now the drain potential is raised, the level of inversion there will decrease and, eventually, strong inversion at that point will cease. For the present, we assume that the drain potential is sufficiently low so that this does *not* occur. With both channel ends strongly inverted, the corresponding surface potentials at these ends will be, from (3.4.6), using $V_{CB} = V_{SB}$ and $V_{CB} = V_{DB}$, respectively,

$$\psi_{s0} \approx \phi_0 + V_{SB} \quad (4.5.1a)$$

$$\psi_{sL} \approx \phi_0 + V_{DB} \quad (4.5.1b)$$

The most commonly used value for ϕ_0 is $2\phi_F$. However, this is not accurate, for reasons discussed near the beginning of Sec. 2.5.2. Thus, $\phi_0 = 2\phi_F + \Delta\phi$ should be used; for the *uniform substrates* we are assuming here, $\Delta\phi = 6\phi_i$ is a good compromise value for ϕ_0 , for common ranges of substrate doping concentration and oxide thickness.

Strong inversion at both ends ensures strong inversion throughout the channel since the surface potential varies monotonically from ψ_{s0} at the source to ψ_{sL} at the drain (Fig. 4.8). As established in Sec. 4.3.1, in strong inversion the current is almost totally due to drift. Thus, we use (4.3.16) to model it, with ψ_{s0} and ψ_{sL} as in (4.5.1).

Denoting by I_{DSN} the current in the case we are presently considering (*both* channel ends strongly inverted), we have then the classical model^{8,14}

$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \left\{ (V_{GB} - V_{FB})(V_{DB} - V_{SB}) - \frac{1}{2} [(V_{DB} + \phi_0)^2 - (V_{SB} + \phi_0)^2] - \frac{2}{3} \gamma [(\phi_0 + V_{DB})^{3/2} - (\phi_0 + V_{SB})^{3/2}] \right\} \quad (4.5.2a)$$

or, after some manipulations,

$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \left\{ (V_{GB} - V_{FB} - \phi_0)(V_{DB} - V_{SB}) - \frac{1}{2} (V_{DB}^2 - V_{SB}^2) - \frac{2}{3} \gamma [(\phi_0 + V_{DB})^{3/2} - (\phi_0 + V_{SB})^{3/2}] \right\} \quad (4.5.2b)$$

Thus, the drain current becomes an *explicit* function of the terminal voltages, a very desirable result, which is to be contrasted with the general models of Sec. 4.3.1. Note also that I_{DSN} is of the form²³

$$I_{DSN} = \frac{W}{L} [g(V_{GB}, V_{DB}) - g(V_{GB}, V_{SB})] \quad (4.5.3)$$

which shows the symmetry of source and drain.† Another form of (4.5.2) is considered in Prob. 4.8. Equation 4.5.2 is the basis of the “level 2” model in the Berkeley Spice simulator.⁷³

DERIVATION DIRECTLY IN STRONG INVERSION. In classical treatments, (4.5.2) is derived directly^{8,14} (rather than from the general case of Sec. 4.3.1), as follows. For any point x in the channel, where the surface potential is $\psi_s(x)$, a quantity $V_{CB}(x)$ is defined such that

$$\psi_s(x) = \phi_0 + V_{CB}(x) \quad (4.5.4)$$

Thus, using (4.5.1), we obtain

$$V_{CB}(0) = V_{SB} \quad (4.5.5a)$$

$$V_{CB}(L) = V_{DB} \quad (4.5.5b)$$

†Expressions (4.5.2) can be written in terms of source referenced voltages by using in it $V_{GB} = V_{GS} + V_{SB}$ and $V_{DB} = V_{DS} + V_{SB}$. This, however, destroys the symmetry in the equations. We will postpone our discussion of source-referenced models until Sec. 4.5.3.

Recalling now the view of the *strong*-inversion layer as an n^+ region, which along with the substrate forms a field-induced n^+p junction (see Sec. 3.3), an easy interpretation becomes possible for $V_{CB}(x)$. It can be viewed as the "effective reverse bias" between the inversion layer at point x and the substrate, changing from V_{SB} at the source to V_{DB} at the drain. We have used the symbol V_{CB} so that the equations of Sec. 3.4.2 become directly applicable here.

Since ϕ_0 is assumed constant, from (4.5.4) we have $d\psi_s(x)/dx = dV_{CB}(x)/dx$. Thus, the drain current, assuming it is due only to drift, can be written from (4.3.3) as follows:†

$$I_{DSN} = \mu W(-Q'_I) \frac{dV_{CB}}{dx} \quad (4.5.6)$$

Integrating from $x = 0$ (where $V_{CB} = V_{SB}$) to $x = L$ (where $V_{CB} = V_{DB}$) produces‡

$$I_{DSN} = \frac{W}{L} \int_{V_{SB}}^{V_{DB}} \mu(-Q'_I) dV_{CB} \quad (4.5.7)$$

The expression for Q'_I can be obtained by using (4.5.4) in (4.3.13):

$$Q'_I = -C'_{ox} \left(V_{GB} - V_{FB} - \phi_0 - V_{CB} + \frac{Q'_B}{C'_{ox}} \right) \quad (4.5.8)$$

where Q'_B is, from (4.3.14) and (4.5.4),

$$Q'_B = -\gamma C'_{ox} \sqrt{\phi_0 + V_{CB}} \quad (4.5.9)$$

†With V_{CB} defined as in (4.5.4), (4.5.6) gives only the drift current and is thus not valid in weak or moderate inversion. However a more general quantity V can be defined such that, when replaced for V_{CB} in (4.5.6), that equation gives the total current (drift plus diffusion)¹⁰ and thus becomes valid in all three regions of inversion. The quantity V is in general different from V_{CB} as defined in (4.5.4) and tends to the latter only in strong inversion (in an energy band context, V is the potential corresponding to the difference between the electron quasi-Fermi level in the inversion layer and the hole quasi-Fermi level in the bulk). This approach is common in the literature.^{10,40-46} Its details will be left for Appendix J, since we have already been able to derive general expressions including both drift and diffusion without having to employ this approach, in the simple manner presented in Sec. 4.3.

‡In some treatments a variable $V_{CS}(x) = V_{CB}(x) - V_{SB}$ is defined. It is easily seen that $V_{CS}(x)$ is equal to the potential at point x in the inversion layer with respect to the source end of the channel. Then (4.5.6) and (4.5.7) become

$$I_{DSN} = \mu W(-Q'_I) \frac{dV_{CS}}{dx} \quad (4.5.6a)$$

$$I_{DSN} = \frac{W}{L} \int_0^{V_{DS}} \mu(-Q'_I) dV_{CS} \quad (4.5.7a)$$

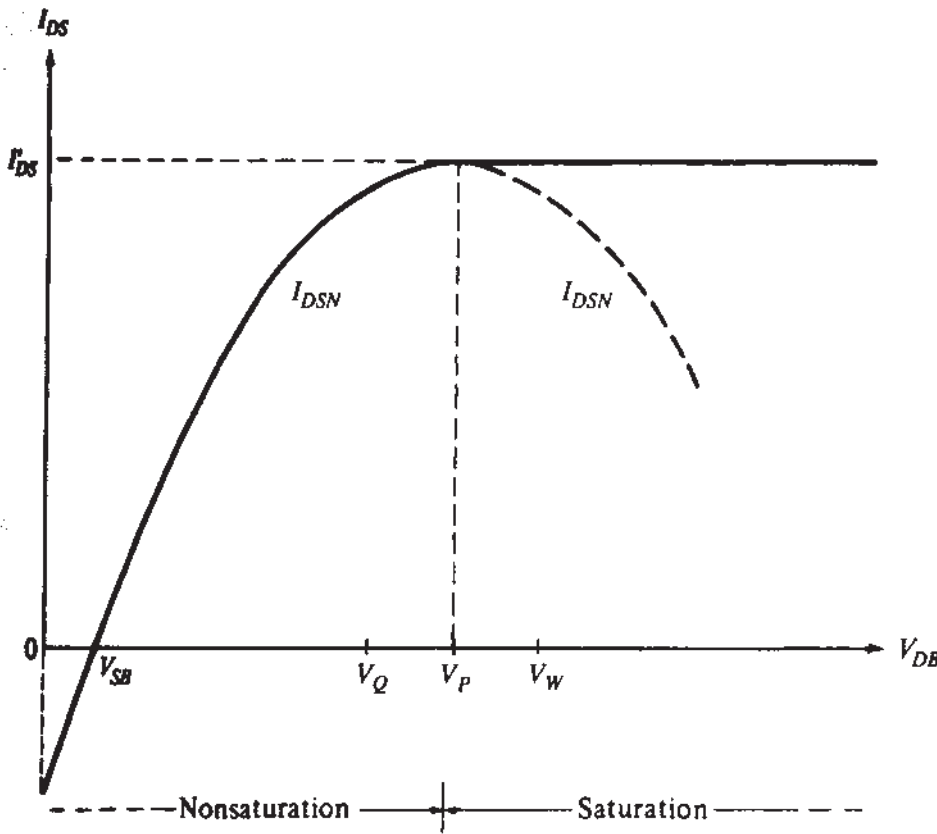


FIGURE 4.14

I_{DSN} computed from (4.5.2), and extension of its maximum value to the range $V_{DB} > V_P$. The solid curves taken together represent the complete strong-inversion model of (4.5.13).

Using this in (4.5.8) gives

$$Q'_I = -C'_{ox} (V_{GB} - V_{FB} - \phi_0 - V_{CB} - \gamma \sqrt{\phi_0 + V_{CB}}) \quad (4.5.10a)$$

$$= -C'_{ox} [V_{GB} - V_{TB}(V_{CB})] \quad (4.5.10b)$$

where $V_{TB}(V_{CB})$ is the gate-substrate extrapolated threshold voltage for an effective reverse bias V_{CB} , as given by (3.4.12c). Using (4.5.10a) in (4.5.7) and integrating gives exactly (4.5.2) again, assuming μ is constant along the channel. (The more general case where μ varies along the channel is discussed in Sec. 4.10.)

A graphical interpretation of I_{DSN} based on (4.5.7) and (4.5.10b) is examined elsewhere.^{14,74,75}

Equation (4.5.2) is plotted in Fig. 4.14 for the same V_{GB} and V_{SB} values as Fig. 4.6. We have extended the plot to V_{DB} values smaller than V_{SB} to show that the current equation is valid for such values. In fact, the smaller the value of V_{DB} , the stronger the level of inversion at the drain end. Thus, the channel remains strongly inverted everywhere for such values, and our development remains valid. Note that V_{DB} is not used below 0, as then the drain-substrate junction would become forward-biased.†

†It should be noted, though, that the models we are presenting work even with V_{SB} and/or V_{DB} slightly negative, provided that the junction currents are negligible.

FORWARD SATURATION. Assume now that $V_{DB} > V_{SB}$, and consider increasing values of V_{DB} . On the horizontal axis in Fig. 4.14 we show the values of V_{DB} from Fig. 4.6. Strictly speaking, (4.5.2) is valid *only* for $V_{DB} < V_Q$. Above V_Q the channel is not in strong inversion near the drain, and thus that equation is not valid. Sometimes, nevertheless, (4.5.2) is used up to the point where the slope of the curve becomes zero; the resulting error may be tolerable in some applications. By setting $dI_{DS}/dV_{DB} = 0$ from (4.5.2), we find that the slope becomes 0 at the following V_{DB} value:

$$V_P = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right)^2 - \phi_0 \quad (4.5.11)$$

The value of the drain current predicted by (4.5.2) with $V_{DB} = V_P$ is denoted by I'_{DS} , as shown in Fig. 4.14. Thus,

$$I'_{DS} = I_{DSN} \big|_{V_{DB}=V_P} \quad (4.5.12)$$

The quantity V_P is the “pinchoff voltage,” which has been encountered in Sec. 3.5.2. According to the derivation and discussion there, it is easy to show that V_P is simply the value of V_{DB} which makes the gate-substrate threshold voltage at the *drain* end of the channel, $V_{TB}(V_{DB})$, equal to the applied gate-substrate voltage V_{GB} .²⁵ It is clear that the value of V_P depends on V_{GB} but has *nothing* to do with V_{SB} , as confirmed by (4.5.11). The value of V_P is close to the quantity V_W in (3.5.2) (which, we remind the reader, is the value of V_{DB} needed to bring the drain end of the channel to the boundary between moderate and weak inversion), as can also be seen in Fig. 4.14. If $\phi_0 = 2\phi_F$ is assumed, V_P reduces to V_W .

As may be concluded from the above derivations, using V_{DB} values up to V_P in (4.5.2) implicitly assumes that the strong-inversion charge expressions are valid for Q'_I arbitrarily close to zero, predicting $Q'_I = 0$ at the drain end of the channel at $V_{DB} = V_P$. (The channel is said to be “pinched off” at that point.) This corresponds to assuming that the broken line in Fig. 3.12, which is a plot of (3.4.11b), actually represents Q'_I all the way down to $Q'_I = 0$, which is clearly incorrect. In addition, if at the drain end of the channel $Q'_I = 0$ at $V_{DB} = V_P$, the carriers would have to travel with infinite drift velocity in order for a nonzero current to be possible, as seen from (1.3.7). (We talk about drift velocity in order to be consistent with the assumption on which the strong inversion model is based, i.e., that all current is due to drift.) To make more physical sense, we will allow for a very small, but nonzero, value of Q'_I at the drain end of the channel in the following discussion; then the carrier speed must be large but finite nevertheless. For $V_{DB} > V_P$, a narrow region is taken to exist between the pinched-off tip of the channel and the drain, with very small $|Q'_I|$ in it; the carriers pass through this region with very high speed. The above region is actually viewed practically as a depletion region, with the excess voltage $V_{DB} - V_P$ dropped across it. The inversion layer then need only support along its length the voltage it supported when $V_{DB} = V_P$. As V_{DB} is increased, the length of the above depletion re-

gion must increase to support the excess voltage, but its length is still assumed very small in comparison to the channel length. Thus, the inversion layer length remains practically at the value L , and, since the voltage across it is still the same as when $V_{DB} = V_P$, the current is still I'_{DS} . We thus have, assuming always that the source end of the channel is strongly inverted,

$$I_{DS} = \begin{cases} I_{DSN}, & V_{DB} \leq V_P \\ I'_{DS}, & V_{DB} > V_P \end{cases} \quad (4.5.13a)$$

$$(4.5.13b)$$

corresponding to the solid curve in Fig. 4.14. Note that I_{DSN} should *not* be used for $V_{DB} > V_P$, because then a completely meaningless behavior is obtained, as shown by the broken curve in the figure. The region $V_{DB} \leq V_P$ is the nonsaturation region and the region for $V_{DB} > V_P$ is the saturation region.[†] Note that despite the fact that in saturation the drain end of the channel is not strongly inverted, strong-inversion theory has been used in this simple model to predict I_{DS} over the whole range covered by (4.5.13). This explains the reason for the common practice of saying that the transistor “operates in strong inversion” for the entire range.

The above explanations concerning the current around $V_{DB} = V_P$ and in saturation are obviously not very satisfying. They have to be used because of the oversimplifying assumptions inherent to the model, as already discussed. Obviously, when this model is used in the transition region between V_Q and V_W or above it, some error can be expected. This will be more apparent in applications where the slope dI_{DS}/dV_{DB} in the above transition region must be known accurately.⁷⁶ Let us recall, after all, that the electric field distribution next to the drain region is two-dimensional, and the “gradual channel approximation” does not hold there in the first place. However, no simple equation exists for the transition region between V_Q and V_W , and so (4.5.13) is often used. A comparison of (4.5.13) to the complete charge sheet model of Sec. 4.3.1 is considered in Prob. 4.11.

If the horizontal depletion region length discussed above cannot be neglected in comparison to L , increasing $V_{DB} - V_P$ will increase the length of that region, and thus it will decrease the “effective” channel length, resulting in non-negligible increase in the drain current. Then (4.5.13) will not be adequate. Such effects become important when L is not too large, and they are treated in Sec. 6.2 where a more careful look at the saturation region is taken.

REVERSE SATURATION. In describing (forward) saturation above, we have assumed $V_{DB} > V_{SB}$. If, instead, $V_{SB} > V_{DB}$ and V_{SB} is raised to V_P and beyond, the phenomena discussed above will take place at the source rather than the drain. This is termed

[†]The nonsaturation region in strong inversion is alternatively called the *triode* region, because the I_{DS} - V_{DS} characteristics in it remind one of the characteristics of a triode electron tube. Similarly, the saturation region in strong inversion is (rarely) referred to as the *pentode* region.

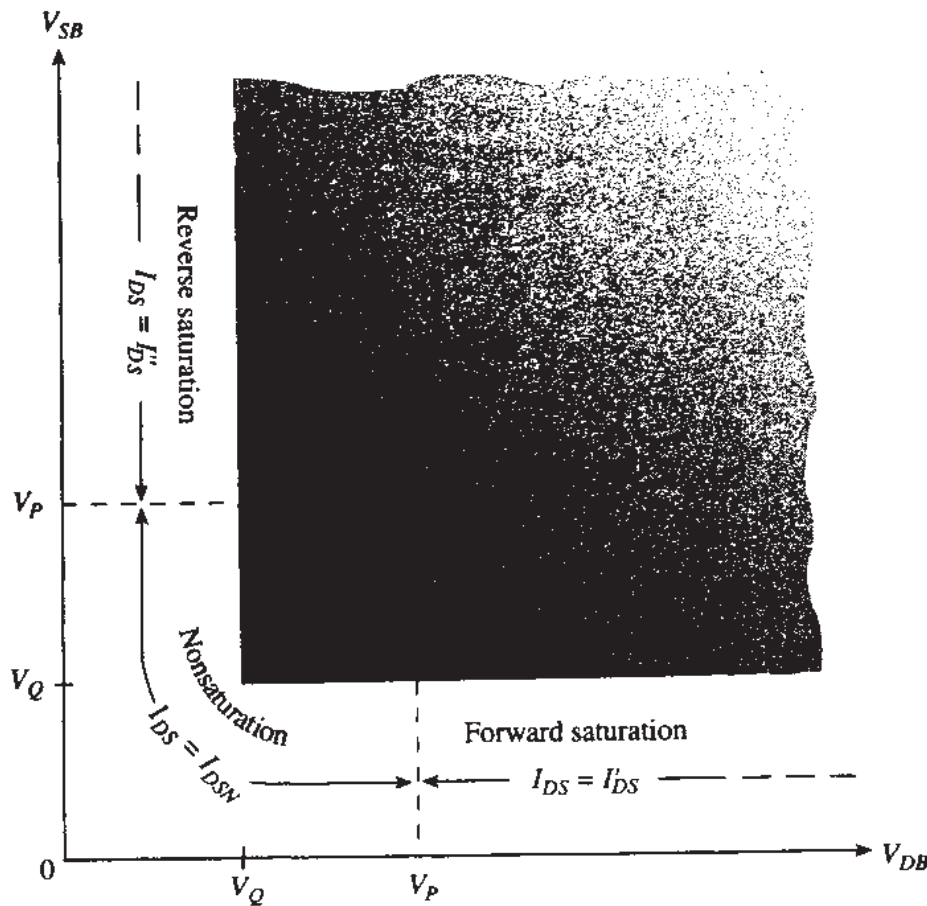


FIGURE 4.15
Complete strong-inversion model.

reverse saturation. In that case, what we have said above will apply to the current entering the *source* terminal with the role of V_{DB} played by V_{SB} , and vice versa. The reverse saturation current, in analogy with the results above, will be given by

$$I_{DS}'' = I_{DSN} \Big|_{V_{SB}=V_P} \quad (4.5.14)$$

COMPLETE STRONG-INVERSION MODEL. Combining the three expressions we developed above (for nonsaturation, forward saturation, and reverse saturation), we have a complete strong-inversion model as shown in Fig. 4.15. Notice that, to guarantee strong inversion at least at one channel end, we keep insisting that at least one of V_{DB} and V_{SB} be smaller than V_Q . In some of the literature this is not done; instead, the only restriction imposed is that one of V_{DB} and V_{SB} be smaller than V_P . With such relaxed assumptions one cannot expect much accuracy of the model we have presented.

4.5.2 Simplified Symmetric Strong-Inversion Model

The above strong-inversion model was seen to result from the complete charge sheet model. Simpler strong inversion expressions will now be seen to result if we start from

the simplified charge sheet models discussed in Sec. 4.3.2. Consider the expression in (4.3.41b). It is clear from the development of that expression that the last term in the sum is the contribution of diffusion currents. Consistent with our practice in the previous subsection, we will neglect this contribution in strong inversion. We thus have

$$I_{DS} = \frac{W}{L} \frac{\mu}{2nC'_{ox}} (Q'_{I0}{}^2 - Q'_{IL}{}^2) \quad (4.5.15)$$

where Q'_{I0} and Q'_{IL} are the inversion charges per unit area at the source and drain ends of the channel, respectively. In nonsaturation, *both* ends are strongly inverted, so we will use for these charges the strong inversion approximation (3.5.14b). This gives the following expression for the *nonsaturation* current:^{68,67,69}

$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \frac{n}{2} [(V_P - V_{SB})^2 - (V_P - V_{DB})^2] \quad (4.5.16)$$

where V_P and n depend on V_{GB} and are given by (3.5.4) and (3.5.11), repeated here for convenience:

$$V_P = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right)^2 - \phi_0 \quad (4.5.17)$$

$$n = 1 + \frac{\gamma}{2\sqrt{\phi_0 + V_P}} \quad (4.5.18)$$

To extend the above results into the saturation region, we follow an approach similar to that of the previous subsection. It can be easily verified that, as V_{DB} is raised, saturation is obtained at $V_{DB} = V_P$, at which point the slope dI_{DSN}/dV_{DB} becomes zero. Using $V_{DB} = V_P$ in (4.5.16), we obtain the value of the forward saturation current as follows:

$$I'_{DS} = \frac{W}{L} \mu C'_{ox} \frac{n}{2} (V_P - V_{SB})^2 \quad (4.5.19)$$

Similarly, if V_{SB} is raised instead, we obtain reverse saturation when $V_{SB} = V_P$, and the reverse saturation current is

$$I''_{DS} = -\frac{W}{L} \mu C'_{ox} \frac{n}{2} (V_P - V_{DB})^2 \quad (4.5.20)$$

The nonsaturation equation (4.5.16) should *not* be used on the saturation regions, as it predicts meaningless results in them. This is because the derivation of this equation relied on the assumption that *both* channel ends were strongly inverted. This assumption is *not* valid in saturation. The above results can be combined into one strong inversion model as in Fig. 4.15.

An even simpler model can be derived as follows. It was seen in Sec. 3.5 that an approximation for V_P is (3.5.12),⁶⁹ repeated below:

$$V_P \approx \frac{V_{GB} - V_{T0}}{n} \quad (4.5.21)$$

where V_{T0} is given by:

$$V_{T0} = V_{FB} + \phi_0 + \gamma\sqrt{\phi_0} \quad (4.5.22)$$

Using (4.5.21) in (4.5.16) gives a model derived by different means elsewhere:^{77,67,68}

$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \left[(V_{GB} - V_{T0})(V_{DB} - V_{SB}) - \frac{n}{2} (V_{DB}^2 - V_{SB}^2) \right] \quad (4.5.23)$$

By setting $dI_{DSN}/dV_{DB} = 0$, it can be easily checked that forward saturation is reached at $V_{DB} = V_P$, with V_P as given by (4.5.21). Using this value for V_{DB} in (4.5.23), we obtain the forward saturation current as follows:

$$I'_{DS} = \frac{W}{L} \mu C'_{ox} \frac{1}{2n} (V_{GB} - V_{T0} - nV_{SB})^2 \quad (4.5.24)$$

Similarly, if V_{SB} is raised instead, reverse saturation is attained at $V_{SB} = V_P$, and the reverse saturation current is

$$I''_{DS} = -\frac{W}{L} \mu C'_{ox} \frac{1}{2n} (V_{GB} - V_{T0} - nV_{DB})^2 \quad (4.5.25)$$

Again, the complete model is as in Fig. 4.15.

The reader may wonder whether, following the several approximations that had to be used in order to derive the above simple models, these models provide satisfactory accuracy. The answer depends on how the parameters in the equations are chosen. If the values dictated from physics are used in them [e.g., if ϕ_F is calculated from (1.4.2)], the accuracy these models provide may not be adequate. If, instead, an optimizer is used to vary these parameters, as if they were free variables, until minimum I_{DS} error is obtained, the accuracy may be improved. This is often done during "parameter extraction," a process considered in Chap. 10. However, this practice is not always without penalty. Choosing a parameter value in order to minimize the error in one quantity in one region can sometimes increase the error of the same quantity in other regions, or of other quantities (e.g., capacitances). We will have more to say about this problem in Chap. 10.

4.5.3 Simplified, Source-Referenced, Strong-Inversion Model

The model discussed in Sec. 4.5.1 resulted from substituting the strong inversion surface potential at the source and drain ends of the channel, in the drift component of the complete charge sheet model.

A simpler strong-inversion model²⁴ can be developed† by substituting $\psi_{s0} = \phi_0 + V_{SB}$ and $\psi_{sL} = \phi_0 + V_{DB}$ from (4.5.1) into the expression for the drift component of the *simplified* source-referenced charge sheet model, (4.3.34). This gives

$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \left[(V_{GB} - V_{SB} - V_{FB} - \phi_0 - \gamma \sqrt{\phi_0 + V_{SB}})(V_{DB} - V_{SB}) - \frac{\alpha}{2} (V_{DB} - V_{SB})^2 \right] \quad (4.5.26)$$

where α has nominally the value given by (4.3.36) with $\psi_{s0} = \phi_0 + V_{SB}$; the value of this quantity will be discussed in detail later.

Substituting $V_{DB} - V_{SB}$ by V_{DS} and $V_{GB} - V_{SB}$ by V_{GS} in the above equation gives

$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \left[(V_{GS} - V_T|_{V_{SB}}) V_{DS} - \frac{\alpha}{2} V_{DS}^2 \right] \quad (4.5.27)$$

where $V_T|_{V_{SB}}$ is the gate-source extrapolated threshold voltage, obtained from (3.4.14b) with $V_{CB} = V_{SB}$.

DERIVATION DIRECTLY IN STRONG INVERSION.²⁴ The complexity of (4.5.2) is caused by the awkward 3/2 powers in it. It is clear that the origin of these powers is the square-root term in (4.5.10a), which, in turn, is due to the expression for Q'_B in (4.5.9). Figure 4.16 shows a plot of $-Q'_B/C'_{ox}$ from that equation. Since the slope of this plot does not vary much, it is reasonable to attempt to approximate $-Q'_B/C'_{ox}$ by the first two terms of its Taylor expansion around the convenient point $V_{CB} = V_{SB}$.‡ This gives

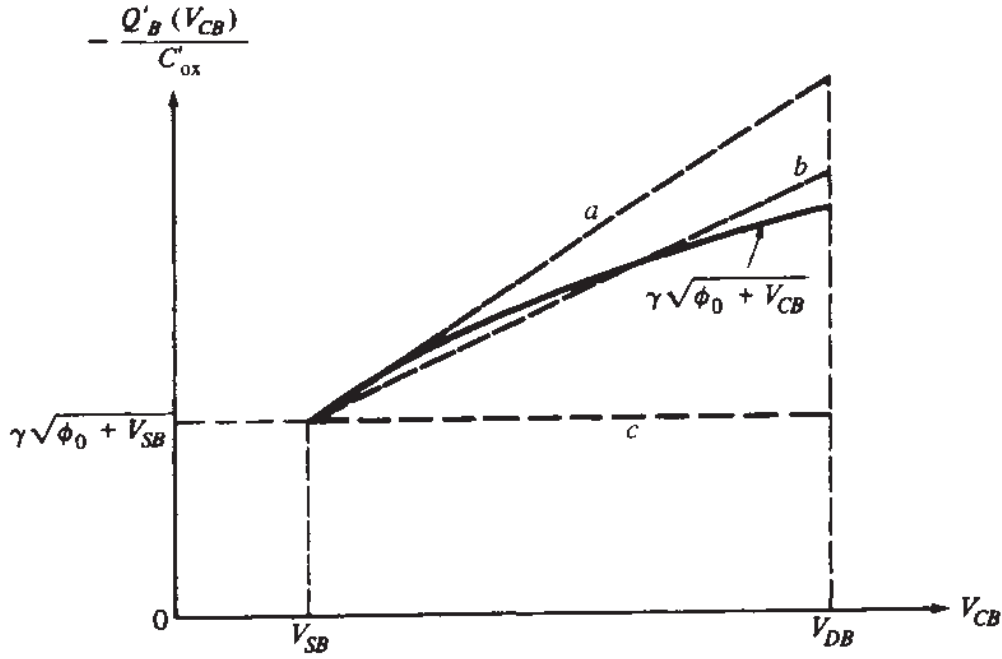
$$-\frac{Q'_B}{C'_{ox}} \approx \gamma \sqrt{\phi_0 + V_{SB}} + (\alpha_1 - 1)(V_{CB} - V_{SB}) \quad (4.5.28)$$

where $\alpha_1 - 1$ is the slope of $-Q'_B/C'_{ox}$ vs. V_{CB} , evaluated at $V_{CB} = V_{SB}$; a value for α_1 will be given later. The right-hand side of (4.5.28) is shown by broken line *a* in Fig. 4.16. It provides the correct value and slope at $V_{CB} = V_{SB}$, but overestimates $-Q'_B/C'_{ox}$ everywhere else. A better approximation can be obtained by lowering the value of the slope by using some value $\alpha < \alpha_1$, as shown by broken line *b*:

$$-\frac{Q'_B}{C'_{ox}} \approx \gamma \sqrt{\phi_0 + V_{SB}} + (\alpha - 1)(V_{CB} - V_{SB}) \quad (4.5.29)$$

†Readers who have skipped Sec. 4.3.2 can go directly to the traditional strong-inversion derivation below.

‡This is equivalent to approximating the threshold voltage along the channel since, from (3.4.14a), $V_T(V_{CB}) = V_{FB} + \phi_0 - Q'_B(V_{CB})/C'_{ox}$.

**FIGURE 4.16**

The quantity $-Q'_B/C'_{ox}$ in strong inversion vs. effective reverse bias in the channel. Solid line: accurate calculation; line *a*: first-order approximation from a Taylor expansion at $V_{CB} = V_{SB}$; line *b*: improved first-order approximation; line *c*: zero-order approximation.

Using Q'_B from (4.5.29) in (4.5.8), we get

$$Q'_i(V_{CB}) = -C'_{ox} [V_{GB} - V_{SB} - V_{FB} - \phi_0 - \gamma\sqrt{\phi_0 + V_{SB}} - \alpha(V_{CB} - V_{SB})] \quad (4.5.30)$$

The drain current can be obtained by using (4.5.30) in (4.5.7) and performing the integration. Substituting $V_{DB} = V_{DS} + V_{SB}$ and $V_{GB} = V_{GS} + V_{SB}$ in the result and assuming μ is constant gives (4.5.27).

In most of the literature, the fact that the threshold voltage is taken at the source is *not* indicated explicitly. For simplicity, we will comply with this practice. Thus, we will write (4.5.27) as

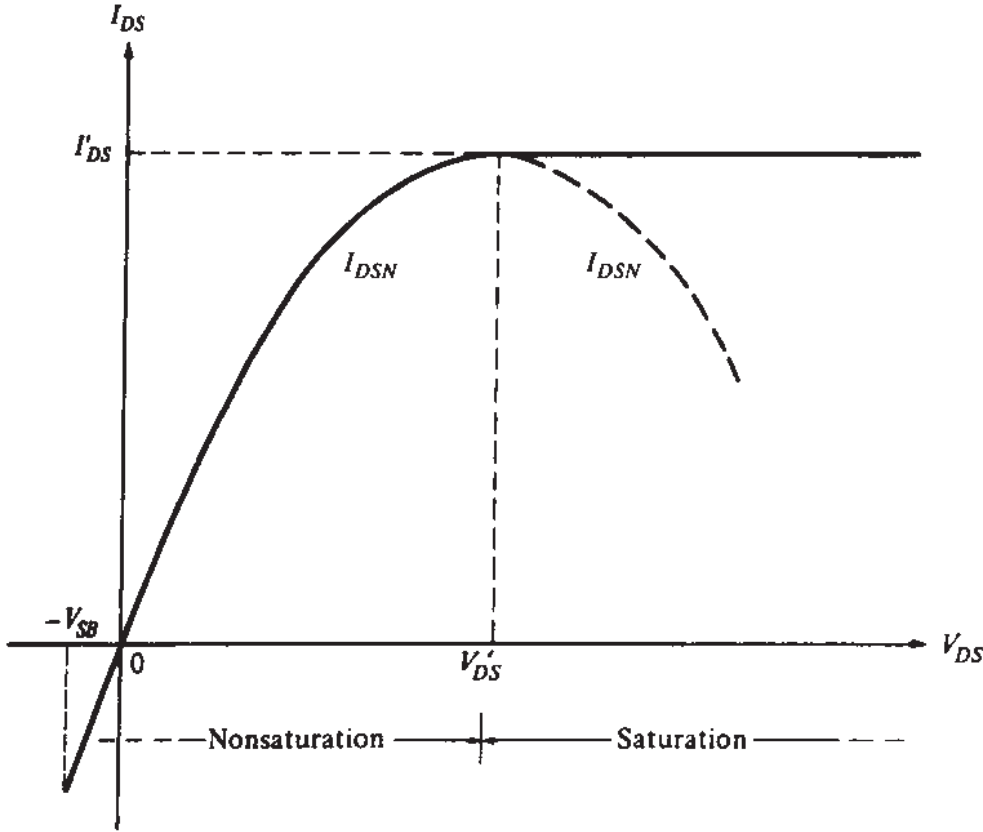
$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \left[(V_{GS} - V_T) V_{DS} - \frac{\alpha}{2} V_{DS}^2 \right] \quad (4.5.31)$$

where, for the rest of this book, V_T will be *defined* as the gate-source extrapolated threshold voltage given by

$$V_T = V_{FB} + \phi_0 + \gamma\sqrt{\phi_0 + V_{SB}} \quad (4.5.32)$$

or

$$V_T = V_{T0} + \gamma(\sqrt{\phi_0 + V_{SB}} - \sqrt{\phi_0}) \quad (4.5.33a)$$

**FIGURE 4.17**

I_{DSN} as computed from (4.5.31), and extension of its maximum value to the range $V_{DS} > V'_{DS}$. The solid curves taken together represent the strong-inversion model in (4.5.37).

with

$$V_{T0} = V_{FB} + \phi_0 + \gamma\sqrt{\phi_0} \quad (4.5.33b)$$

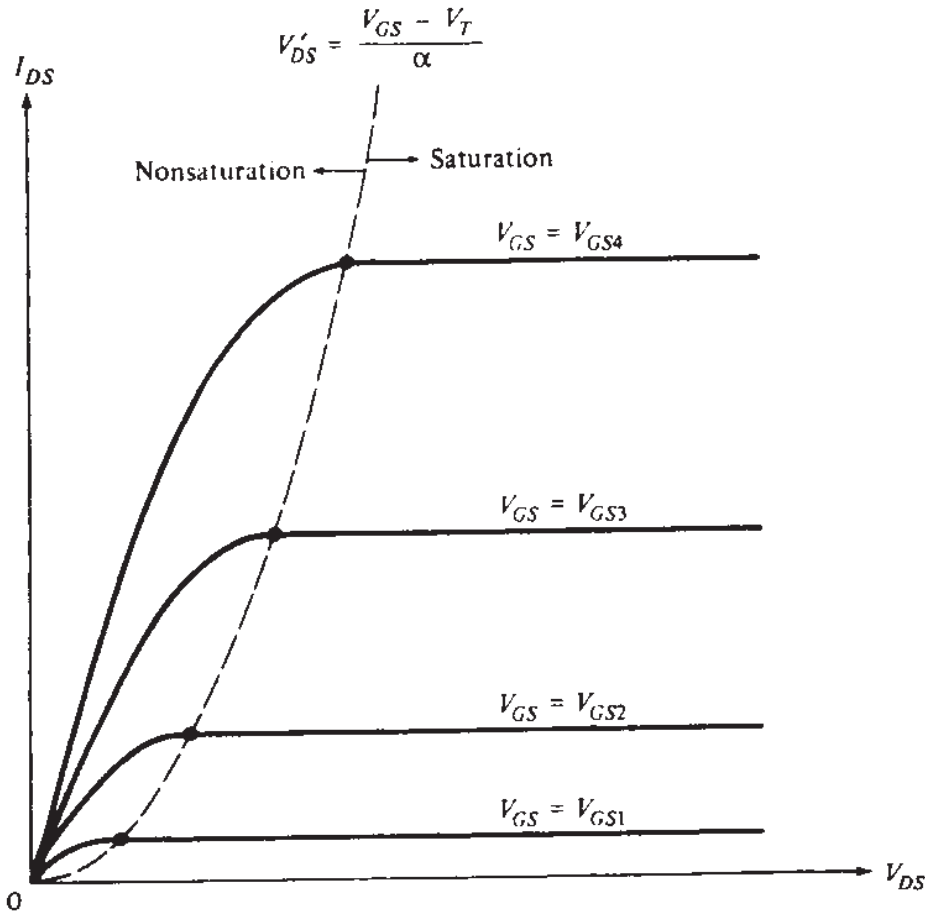
Clearly, V_T depends on V_{SB} owing to the body effect. In fact, in many treatments the term *body effect* implies only the dependence of V_T on V_{SB} . A more general point of view was provided in Sec. 3.3. The parameter ϕ_0 is often taken equal to $2\phi_F$, but this is not justifiable, as has already been explained near the beginning of Sec. 2.5.2.

A plot of I_{DSN} from (4.5.31) is shown in Fig. 4.17. The value V'_{DS} of V_{DS} at which the maximum occurs is found from (4.5.31) by setting $dI_{DSN}/dV_{DS} = 0$ and solving for $V_{DS} = V'_{DS}$:

$$V'_{DS} = \frac{V_{GS} - V_T}{\alpha} \quad (4.5.34)$$

The corresponding value of the drain current I'_{DS} is found by using $V_{DS} = V'_{DS}$ from (4.5.34) in (4.5.31):

$$I'_{DS} = \frac{W}{L} \mu C'_{ox} \frac{(V_{GS} - V_T)^2}{2\alpha} \quad (4.5.35)$$

**FIGURE 4.18**

I_{DS} - V_{DS} characteristics as obtained from the simplified strong-inversion model of (4.5.37).

As in the accurate model, this is taken to be the value of I_{DS} for $V_{DS} > V'_{DS}$ also. The complete model then becomes, in analogy to (4.5.13):

$$I_{DS} = \begin{cases} I_{DSN}, & V_{DS} \leq V'_{DS} \\ I'_{DS}, & V_{DS} > V'_{DS} \end{cases} \quad (4.5.36a)$$

$$(4.5.36b)$$

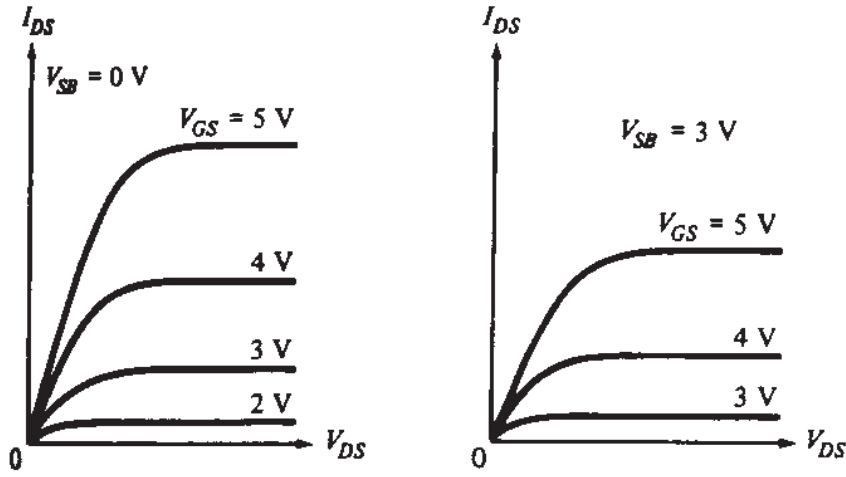
Thus, using the equations developed for I_{DSN} and I'_{DS} , we have

$$I_{DS} = \begin{cases} \frac{W}{L} \mu C'_{ox} \left[(V_{GS} - V_T) V_{DS} - \frac{\alpha}{2} V_{DS}^2 \right], & V_{DS} \leq V'_{DS} \\ \frac{W}{L} \mu C'_{ox} \frac{(V_{GS} - V_T)^2}{2\alpha}, & V_{DS} > V'_{DS} \end{cases} \quad (4.5.37a)$$

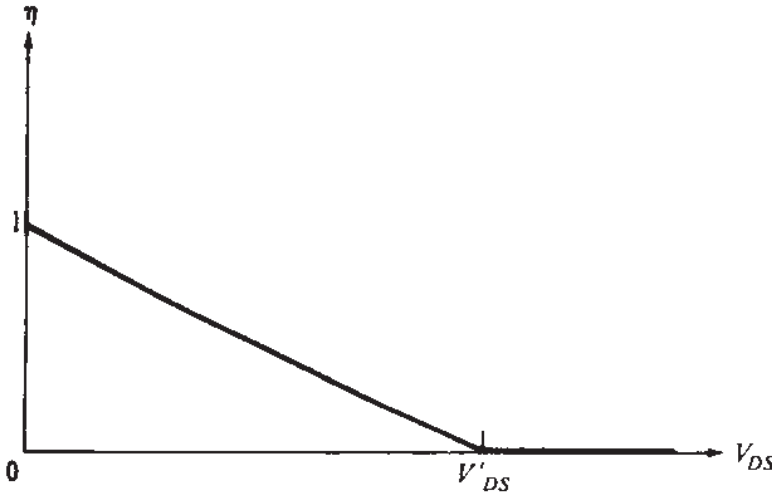
$$(4.5.37b)$$

This corresponds to the solid line in Fig. 4.17. A set of I_{DS} - V_{DS} characteristics for various V_{GS} values is shown in Fig. 4.18. Such characteristics depend on V_{SB} through V_T in (4.5.33). This is illustrated in Fig. 4.19.

The above model²⁴ is the basis of the "level 3" model in the Berkeley Spice simulator.⁷³

**FIGURE 4.19**

I_{DS} - V_{DS} characteristics for two different V_{SB} values.

**FIGURE 4.20**

Parameter η vs. V_{DS} .

Equation (4.5.37) can be put in a very compact form by defining a convenient parameter η as follows:^{78,79}

$$\eta = \begin{cases} 1 - \frac{V_{DS}}{V'_{DS}}, & V_{DS} \leq V'_{DS} \end{cases} \quad (4.5.38a)$$

$$\begin{cases} 0, & V_{DS} > V'_{DS} \end{cases} \quad (4.5.38b)$$

This parameter is plotted in Fig. 4.20. The drain current can now be expressed as follows:

$$I_{DS} = I'_{DS} (1 - \eta^2), \quad \text{both nonsaturation and saturation} \quad (4.5.39)$$

where I'_{DS} is given by (4.5.35).

It is easy to verify that at $V_{DS} = V'_{DS}$ (4.5.30) predicts $Q'_l = 0$ at the drain end of the channel. The reason for this unrealistic result is as discussed for the accurate

strong-inversion model, and will not be repeated. Similarly, the comments made for that model concerning possible inaccuracies in the neighborhood of $V_{DS} = V'_{DS}$ apply here as well. Special "smoothing functions" are often used to improve the transition from nonsaturation to saturation.⁸⁰⁻⁸²

We now return to the problem of choosing an appropriate value for α in (4.5.29) and, therefore, in (4.5.37). The derivation of MOSFET characteristics in the early days implicitly assumed the following value for α :

$$\alpha_0 = 1 \quad (4.5.40)$$

This corresponds to line *c* in Fig. 4.16, which is a very poor approximation to the solid line in the same figure. This approximation is equivalent to assuming that the depletion region depth is the same all along the channel and equal to its actual value at the source. The result is that $|Q'_B|$ is underestimated everywhere except at the source. This, from (4.5.8), results in an overestimation of $|Q'_I|$ and a resulting overestimation of I_{DS} . The error in I_{DS} can be *large* for devices in which γ is not small; we will come back to this point later. Note also that, since $|Q'_I|$ is overestimated, an artificially large value of V_{DS} would be needed to reduce $|Q'_I|$ to zero at the drain. This value of V_{DS} is V'_{DS} , as already mentioned, under the simplified assumptions in the present model. The overestimation of V'_{DS} is also seen from (4.5.34) with $\alpha = 1$, rather than with a more realistic value of α larger than unity.

The next value we will consider for α is the one that results from the Taylor expansion leading to (4.5.28).²⁴ This value corresponds to line *a* in Fig. 4.16 and is given by

$$\alpha_1 = 1 + \frac{\gamma}{2\sqrt{\phi_0 + V_{SB}}} \quad (4.5.41)$$

It is clear from the figure that this would be a good value to use for α only if $V_{DS} = V_{DB} - V_{SB}$ were small. In the general case, though, it results in overestimating $|Q'_B|$ and, hence, underestimating $|Q'_I|$. Thus, I_D and V'_{DS} will be underestimated, i.e., the errors here are in the opposite direction from those for $\alpha = \alpha_0$. Clearly, one should seek values for α between α_0 and α_1 , resulting in the approximation indicated by line *b* in Fig. 4.16. Several such values have been proposed in the literature, and they usually correspond to modifications of (4.5.41) to lower its value. One such modification has the form

$$\alpha_2 = 1 + d_2 \frac{\gamma}{2\sqrt{\phi_0 + V_{SB}}} \quad (4.5.42)$$

where d_2 is a correction factor. Values used^{24,83} for d_2 vary between 0.5 and 0.8. In another approach, this correction factor is allowed to depend on V_{SB} for better fit,⁸⁴

†It is interesting to note here that, as can be verified by using (4.5.32), we have

$$\alpha_1 = 1 + \frac{dV_T}{dV_{SB}}$$

and the expression $d_2 = 1 - [k_1 + k_2(\phi_0 + V_{SB})]^{-1}$ is suggested, with k_1 and k_2 being constants chosen to minimize overall error.[†] This, however, detracts from the simplicity of the model.

Another value suggested for α is⁸⁵

$$\alpha_3 = 1 + \frac{\gamma}{2\sqrt{\phi_3 + \phi_0 + V_{SB}}} \quad (4.5.43)$$

where $\phi_3 = 1$ V. This is a good empirical expression, both for its accuracy and for its simplicity, and gives satisfactory results for practical situations.

Finally, one can seek a zero-order estimate for α for rough calculations, which is even independent of V_{SB} . One such estimate is⁸⁶

$$\alpha_4 = 1 + \frac{\gamma}{4\sqrt{\phi_0}} \quad (4.5.44)$$

All the above three expressions for α have been arrived at with the goal of providing satisfactory accuracy over a reasonable range of bias voltages in strong inversion. One should not be very surprised if they do not work very well in extreme cases (e.g., for V_{DS} in excess of 10 V). However, they do provide satisfactory accuracy for most practical cases. In general, which expression is best to use for α depends on the accuracy sought, the desired speed of calculations, and the range of bias voltages expected. The final criterion is how well the model predicts the drain current. One cannot judge this easily by looking at the error in predicting Q'_B . Of course the errors in Q'_B and I_{DS} are related, but not in a very direct manner. For example, one can have a large relative error in Q'_B and still have a small error in I_{DS} if Q'_B is small in the first place [so that $|Q'_B/C'_{ox}|$ is small in comparison to the other terms in (4.5.8)]. This will be the case for devices with thin oxides and lightly doped substrates; then γ is small, leading to a small value of $|Q'_B|$ in (4.5.9a).

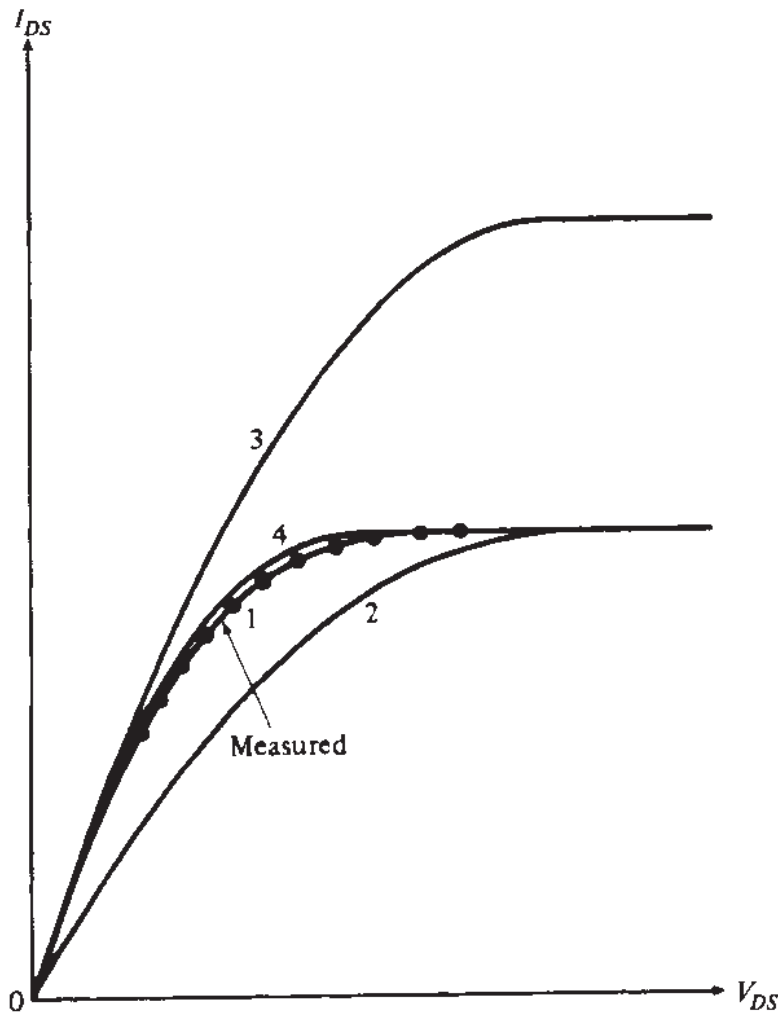
Since the final criterion of success is how well I_{DS} is predicted by the approximate model, one could have considered deriving (4.5.27) directly from the accurate current expression in (4.5.2). This is indeed possible but amounts to little more than a mathematical exercise, not revealing the underlying physical assumptions in the approximate model. The approach is considered in Prob. 4.12.

A comparison of the various values of α is suggested in Prob. 4.14.

Note that when γ is *small*, the values of α in (4.5.41) to (4.5.44) are all close to 1; i.e., the value in (4.5.40) is then reasonable. In this case, (4.5.37) becomes

$$I_{DS} = \begin{cases} \frac{W}{L} \mu C'_{ox} \left[(V_{GS} - V_T) V_{DS} - \frac{1}{2} V_{DS}^2 \right], & V_{DS} \leq V'_{DS}, \text{ small } \gamma \quad (4.5.45a) \\ \frac{W}{L} \mu C'_{ox} \frac{(V_{GS} - V_T)^2}{2}, & V_{DS} > V'_{DS}, \text{ small } \gamma \quad (4.5.45b) \end{cases}$$

[†]The values of k_1 and k_2 depend on the error criterion chosen and the range of voltages over which small error is desired. Values used^{84,87,88} include $k_1 = 1.744$ with $k_2 = 0.8364 \text{ V}^{-1}$, and $k_1 = 1.41$ with $k_2 = 0.43 \text{ V}^{-1}$.

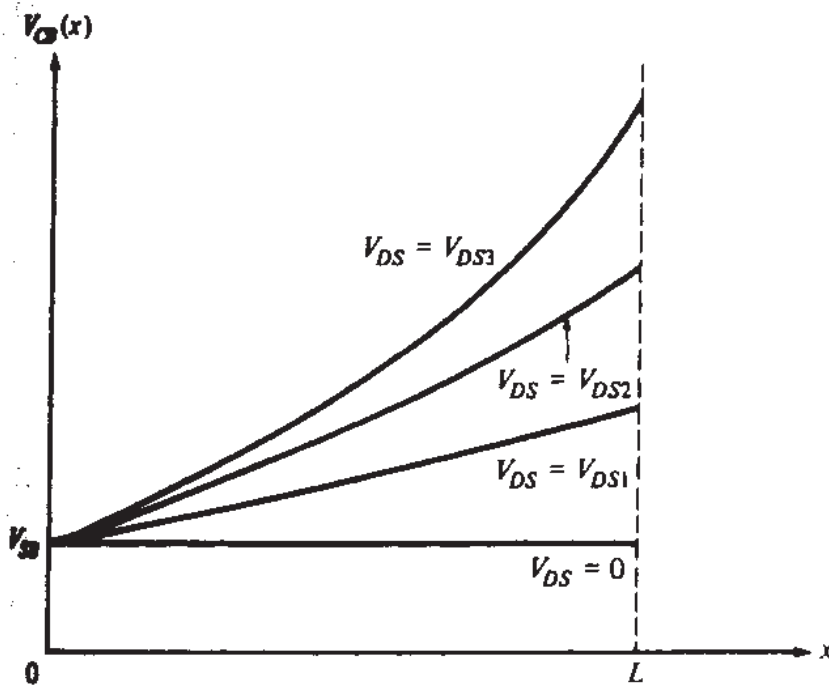
**FIGURE 4.21**

I_{DS} vs. V_{DS} for fixed V_{SB} and $V_{GS} = 5$ V. Curve 1: measured characteristic of n -channel transistor on a heavily doped substrate (taken from a 14007-type CMOS inverter); curve 2: approximate model with $\alpha = 1$, adjusted for good matching in the saturation region; curve 3: approximate model with $\alpha = 1$, adjusted for good matching at very low V_{DS} values; curve 4: approximate model with $\alpha = 1.7$, chosen for good matching at very low V_{DS} and in the saturation region.

where $V'_{DS} = V_{GS} - V_T$. These equations have been used widely for approximate hand calculations for circuit design, and even for quick computer calculations,²¹ and are the basis of the "level 1" model in Spice.⁷³ However, in circuits literature these equations are often used indiscriminately even for devices with large γ , which can result in *serious* error. As an example, consider the measured characteristics shown in Fig. 4.21 by curve 1. The device used was chosen to have a large γ , in order to prove our point. Curve 2 represents (4.5.45), with the value for $(W/L)\mu C'_{ox}$ adjusted for good matching in saturation. The model clearly fails in nonsaturation, and it predicts a V'_{DS} which is too large. If $(W/L)\mu C'_{ox}$ is instead adjusted for a good fit in the low nonsaturation region, then the model gets out of hand in saturation, as shown by curve 3. If the model of (4.5.37) is used instead, we have the additional flexibility of choosing α ; with $\alpha = 1.7$, we obtain curve 4.

For *very* crude modeling work, the strong-inversion equations are sometimes taken to be valid for V_{GS} as low as V_T , where they predict $I_{DS} = 0$. In such cases, the moderate- and weak-inversion regions are not considered.

POTENTIAL VS. POSITION. In strong inversion, one can develop a rather simple relation between the position x in the channel and the effective reverse bias $V_{CB}(x)$ at that point, extending an idea developed in Sec. 4.3.1. All strong-inversion expressions we have given for the nonsaturation current can be written in the form

**FIGURE 4.22**

Channel "effective reverse bias" with respect to the substrate vs. distance from the source for the simplified strong-inversion model and for four V_{DS} values.

$$I_{DSN} = \frac{W}{L} h(V_{GB}, V_{SB}, V_{DB}) \quad (4.5.46)$$

with the function h depending on which model is being used. If we consider point x in the channel as the drain of a fictitious transistor with length x , we will have

$$I_{DSN} = \frac{W}{x} h(V_{GB}, V_{SB}, V_{CB}(x)) \quad (4.5.47)$$

Eliminating I_{DSN} among the two equations gives

$$\frac{x}{L} = \frac{h(V_{GB}, V_{SB}, V_{CB}(x))}{h(V_{GB}, V_{SB}, V_{DB})} \quad (4.5.48)$$

This equation gives the relation between x and $V_{CB}(x)$. As an example, using the approximate model developed above, results in a simple expression. Extending this expression to include the saturation region, and using the parameter η in (4.5.38), we can put it in the form (Prob. 4.17)

$$V_{CB}(x) = V_{SB} + \frac{V_{GS} - V_T}{\alpha} \left[1 - \sqrt{1 - \frac{x}{L} (1 - \eta^2)} \right] \quad (4.5.49)$$

This relation is plotted for various values of V_{DS} in Fig. 4.22 ($0 < V_{DS1} < V_{DS2} < V_{DS3}$). As seen, for small V_{DS} values V_{CB} is nearly proportional to x . This is because

for such V_{DS} values the inversion layer is nearly uniform, so the voltage is distributed almost evenly along its length. At larger V_{DS} values, this is no longer the case. This can be explained easily. For a given V_{DS} , say $V_{DS} = V_{DS3}$, one can find from the plot the voltage drop across a chunk of the inversion layer of length Δx between two points x_1 and x_2 . The drop will simply be $\Delta V_{CB} = V_{CB}(x_2) - V_{CB}(x_1)$. If another chunk of the same length Δx is taken closer to the drain, the voltage ΔV_{CB} across it will be larger, for the following reason. The second chunk contains a smaller inversion layer charge per unit area, since the surface potential increases, and thus $|Q'_I|$ decreases, as one goes from the source toward the drain [this follows, for example, from (4.5.30)]. Thus a larger voltage will be needed across the second chunk to support the same current (Sec. 1.3.2). From this argument, it is clear why the slope dV_{CB}/dx increases with increasing x . A related observation concerns the drift velocity of electrons. As follows from (1.3.7), to carry the same current the electrons must move faster at points where the inversion charge magnitude per unit area is smaller. Since the electrons are in abundance at the source end of the channel, they can carry the required charge per unit time (current) while moving slowly. As they approach the drain, they gain speed. Thus, in a given small element of length Δx near the drain, fewer electrons are to be found at any given instant than within a corresponding element near the source. If V_{DS} is large, the electrons can be traveling so fast that "velocity saturation" can occur near the drain. The effect of this on the transistor I - V characteristics can become significant in short-channel devices, and is considered in Chap. 6.

If $V_{CB}(x)$ were plotted for $V_{DS} = V'_{DS}$ using (4.5.49), the slope would be infinite at $x = L$. This physically impossible result, corresponding to an infinite electric field at that point, is simply a result of the limitations of the model as $V_{DS} = V'_{DS}$ is approached. As already discussed, the model implicitly assumes $Q'_I = 0$ at the drain for $V_{DS} = V'_{DS}$; this is wrong as already explained. In addition, the model fails to take into account the presence of diffusion currents. Whereas, in strong inversion, the I_{DS} component due to diffusion is small, the portion of the channel current that is due to diffusion, $I_{diff}(x)$ in (4.3.4), increases with increasing x at the expense of $I_{drift}(x)$, as has already been remarked.⁴⁹ Thus, any conclusion reached at $V_{DS} = V'_{DS}$, based on the assumption of drift current only, is subject to question.

COMPARISON TO THE COMPLETE STRONG-INVERSION MODEL. In strong inversion, if the complete strong inversion model introduced in Sec. 4.5.1 is compared to the complete charge sheet model of Sec. 4.3.1, the agreement is found to be very good. Compared to these models, the simplified strong-inversion model of Sec. 4.5.3 is somewhat in error (e.g., up to 5 percent or more in terms of current values). In practice, though, several factors contribute to the simplified model being favored over the complete one in a variety of situations. These factors are summarized below:

1. The simplified strong inversion model is simple. This is a desirable feature when very large circuits must be analyzed by computer or when quick hand calculations are needed.
2. Practical devices exhibit high-order effects which are not taken into account by either model; for example, the substrate is never exactly uniform, owing to fabrica-

tion details (Chap. 5). This tends to make the difference between the two models less significant, in terms of accuracy.

3. The assumptions behind each model will be seen later to lead to corresponding expressions for the total charges (Chap. 7) and the capacitances (Chap. 8) of the device. In the case of the simplified strong-inversion model, the expressions for these quantities are simple. In contrast to this, the expressions for some of these quantities, as derived for the complete strong-inversion model, are so complicated as to be totally impractical. In addition, the assumptions behind the simplified model are known to lead to simple models suitable for very high-frequency operation (Chap. 9); this is not known to be possible for the complete model.
4. The simplified model contains explicitly the threshold voltage V_T . This is an important parameter that is widely discussed, used, and measured.[†] In fact, even for short-channel and/or narrow-channel devices, use is made of the simplified model with V_T replaced by an effective threshold (Chap. 6). A large amount of work has been done on how to estimate and measure this quantity in such cases. In contrast, the complete model does not contain the threshold explicitly, and it is more difficult to adopt it for use with short- and/or narrow-channel devices.

Thus, the more versatile simplified model is favored for general use in strong inversion. In cases, however, where the nuances of strong inversion need to be studied, the complete model is to be preferred.

RELATION TO SYMMETRIC MODEL. The source-referenced model we have discussed in this subsection is related to the symmetric models of Sec. 4.5.2. To see this, in (4.5.23) use $V_{GB} = V_{GS} + V_{SB}$ and $V_{DB} = V_{DS} + V_{SB}$. Then the equation can be manipulated to take the form of (4.5.37a), with $V_T = V_{T0} + (n - 1)V_{SB}$ [in lieu of (4.5.33)] and with n in lieu of α (Prob. 4.16).

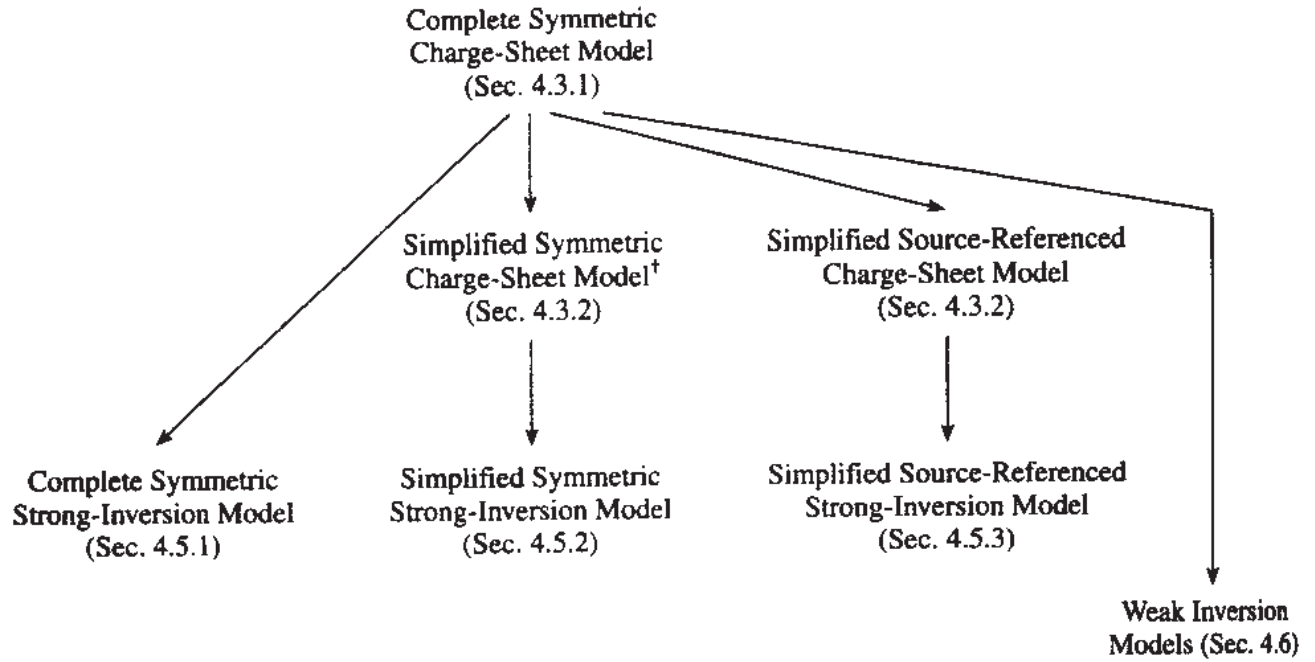
4.5.4 Model Origin Summary

Several models have already been discussed. To help place these models, and their conceptual origin, as derived in this book, in perspective, we provide the chart of Fig. 4.23.[‡] Note that the term *charge sheet model*, if no specific region of inversion is mentioned, commonly implies a model valid in all regions of inversion. All models mentioned in Fig. 4.23 are today in use, and one or the other may be preferable depending on the application.

In Fig. 4.23 we also show models for weak inversion, which are the subject of the next section.

[†]In this book, the term *threshold voltage* will always imply *extrapolated* threshold voltage unless stated otherwise. Unfortunately, in the literature on MOS devices the term "threshold voltage" is used with several different meanings. This is discussed in Sec. 4.15.

[‡]We note that, historically, the derivation of these models did not follow the flow indicated in Fig. 4.23. References to the historical origins of each model have been given in the corresponding sections.



† Includes the charge-based and current-based versions discussed in Sec. 4.3.2.

FIGURE 4.23
Model hierarchy.

4.6 WEAK INVERSION

As seen from the last entry in Table 4.1, for a transistor operating in weak inversion^{26–39} *no* part of the channel is moderately or strongly inverted. The conditions for weak inversion operation have been given in terms of terminal voltages in Sec. 4.4.

The general model developed in Sec. 4.3.1 can be used to produce a simplified equation for I_{DS} in the weak-inversion region. This will be discussed later. For the present, we prefer to show the classical approach for obtaining such an equation to provide some independent intuition about this region of operation.

As follows from the material in Sec. 3.4.3, for a weakly inverted point in the channel, the surface potential satisfies

$$\psi_s \approx \psi_{sa}(V_{GB}) \quad (4.6.1)$$

where

$$\psi_{sa}(V_{GB}) = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right)^2 \quad (4.6.2)$$

Let us assume for the sake of simplicity that ψ_s is *exactly* equal to $\psi_{sa}(V_{GB})$; this will not affect the forthcoming analysis. Since the surface potential depends only on

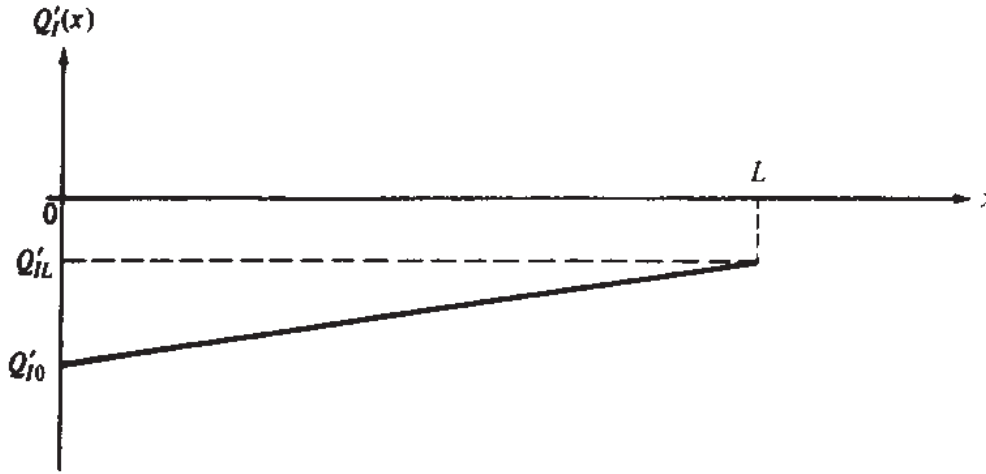


FIGURE 4.24

Inversion layer charge per unit area vs. distance from the source in weak inversion.

V_{GB} , it is independent of the position along the channel. This implies two important facts:

1. Q'_B will be independent of position along the channel, as seen from (4.3.14). This means that the depletion region depth does not change along the channel.
2. Since all points at the surface are assumed at the same potential with respect to the substrate, the potential difference between such points is zero. Therefore, the electric field has a zero horizontal component. If there is current through the channel, then, it cannot be caused by drift; thus, all current must be caused by *diffusion*.

It follows then from the material in Sec. 1.3.3 that the plot of Q'_I vs. x must be a straight line (Fig. 4.24). Thus, (1.3.22) gives, with $Q' = Q'_I$, $b = W$, and $a = L$:

$$I_{DS} = -\frac{W}{L} \mu \phi_t (Q'_{I0} - Q'_{IL}) \quad (4.6.3)$$

Indeed, this is the same as (4.3.12), which was seen to give the contribution of diffusion current to I_{DS} . Thus, (4.6.3) can be considered as resulting directly from the charge sheet model of Sec. 4.3.1.

SYMMETRIC MODEL. The two values of Q'_I in the above equation can be found from (3.4.23), which is valid in weak inversion and even in depletion:

$$Q'_{I0} = -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\psi_{sa}(V_{GB})}} \phi_t e^{[\psi_{sa}(V_{GB}) - 2\phi_F]/\phi_t} \cdot e^{-V_{SB}/\phi_t} \quad (4.6.4)$$

$$Q'_{IL} = -\frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\psi_{sa}(V_{GB})}} \phi_t e^{[\psi_{sa}(V_{GB}) - 2\phi_F]/\phi_t} \cdot e^{-V_{DB}/\phi_t} \quad (4.6.5)$$

Using these in (4.6.3) we obtain

$$I_{DS} = \frac{W}{L} \hat{I}(V_{GB}) (e^{-V_{SB}/\phi_t} - e^{-V_{DB}/\phi_t}) \quad (4.6.6)$$

where

$$\hat{I}(V_{GB}) = \mu \frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\psi_{sa}(V_{GB})}} \phi_t^2 e^{[\psi_{sa}(V_{GB}) - 2\phi_F]/\phi_t} \quad (4.6.7)$$

Readers familiar with bipolar transistors will recognize the similarity of (4.6.6) to the Ebers-Moll equations.⁵⁸ This should not be surprising, as similar mechanisms are responsible for current flow in the bipolar transistor (under common assumptions) and the weakly inverted MOS transistor.

Since (3.4.23), which was used in the above development, is valid even in depletion, (4.6.6) will be valid even if V_{DB} is so large that the drain end of the channel is less than "weakly" inverted.

ALTERNATE EXPRESSIONS. In order to provide continuity with the technical literature, we give here two alternate expressions for I_{DS} .⁶⁶⁻⁶⁹ The first one is obtained from (4.6.6)–(4.6.7), by using (3.4.3), (3.5.9), and (3.5.11):

$$I_{DS} = \frac{W}{L} \mu C'_{ox} (n - 1) e^{(\phi_0 - 2\phi_F)/\phi_t} \phi_t^2 \left[e^{(V_P - V_{SB})/\phi_t} - e^{(V_P - V_{DB})/\phi_t} \right] \quad (4.6.8)$$

The second expression is obtained by using in (4.6.8) the approximation for V_P in (3.5.12). This gives

$$I_{DS} = \frac{W}{L} \mu C'_{ox} (n - 1) e^{(\phi_0 - 2\phi_F)/\phi_t} \phi_t^2 \left[e^{(V_{GB} - V_{T0} - nV_{SB})/(n\phi_t)} - e^{(V_{GB} - V_{T0} - nV_{DB})/(n\phi_t)} \right] \quad (4.6.9)$$

with n as given by (3.5.11).

SOURCE-REFERENCED MODEL. Equation (4.6.3) can be rewritten as follows:

$$I_{DS} = -\frac{W}{L} \mu \phi_t Q'_{I0} \left(1 - \frac{Q'_{IL}}{Q'_{I0}} \right) \quad (4.6.10)$$

From (4.6.4) and (4.6.5) we have

$$\frac{Q'_{IL}}{Q'_{I0}} = e^{-(V_{DB} - V_{SB})/\phi_t} = e^{-V_{DS}/\phi_t} \quad (4.6.11)$$

Thus, I_{DS} becomes

$$I_{DS} = \frac{W}{L} \mu \phi_t (-Q'_{i0}) (1 - e^{-V_{DS}/\phi_t}) \quad (4.6.12)$$

which can be evaluated by using (4.6.4). A compact expression³⁵ can be developed by using the approximation in (3.4.28b), with subscript C replaced by S . Using this in (4.6.12) gives

$$I_{DS} = \frac{W}{L} I'_M e^{(V_{GS}-V_M)/(n\phi_t)} (1 - e^{-V_{DS}/\phi_t}), \quad \text{fixed } V_{SB} = V'_{SB} \quad (4.6.13)$$

where V_M is the upper limit of weak inversion in terms of V_{GS} , for the given value of V_{SB} :

$$V_M = V_{FB} + 2\phi_F + \gamma \sqrt{2\phi_F + V'_{SB}} \quad (4.6.14)$$

and

$$I'_M = \mu \frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{2\phi_F + V'_{SB}}} \phi_t^2 \quad (4.6.15)$$

and with the quantity n evaluated from (3.4.25) at $\psi_{sa} = 2\phi_F + V'_{SB}$ (see Fig. 3.7):

$$n = 1 + \frac{\gamma}{2\sqrt{2\phi_F + V'_{SB}}} \quad (4.6.16)$$

It is emphasized that this equation, along with (4.6.13), is useful for studying the approximate variation of I_{DS} with V_{GS} , when V_{SB} is *fixed* at some value denoted by V'_{SB} . Trying to use this equation to study the variation of I_{DS} with V_{SB} is *not* advisable, since several quantities in the equation (V_M , I'_M , n) depend on V'_{SB} in a complicated manner. In such cases, (4.6.6)–(4.6.7) should be used. It is clear from (4.6.6) that I_{DS} depends on V_{SB} *only* through the quantity $\exp(-V_{SB}/\phi_t)$.

I_D is plotted vs. V_{DS} by using (4.6.13) in Fig. 4.25, with V_{GS} as a parameter, for a fixed V_{SB} . I_D is practically the same as I_{DS} if the latter is much larger than the leakage currents (see below). As seen, the curves become horizontal for V_{DS} larger than a few ϕ_t , since the last exponential in the equation becomes negligible compared to 1. This happens at V_{DS} values which are *independent* of V_{GS} , a fact which is in sharp contrast to the case of strong-inversion operation (Fig. 4.18).

Using equal V_{GS} steps, the vertical spacing of successive curves in Fig. 4.25 for a given V_{DS} increases nearly exponentially. This exponential behavior is brought out clearly by plotting $\log I_D$ vs. V_{GS} with V_{DS} fixed, as shown in Fig. 4.26. The result is nearly a straight line in weak inversion. Above this region we have moderate inversion, where I_{DS} does not vary exponentially with V_{GS} . At the bot-

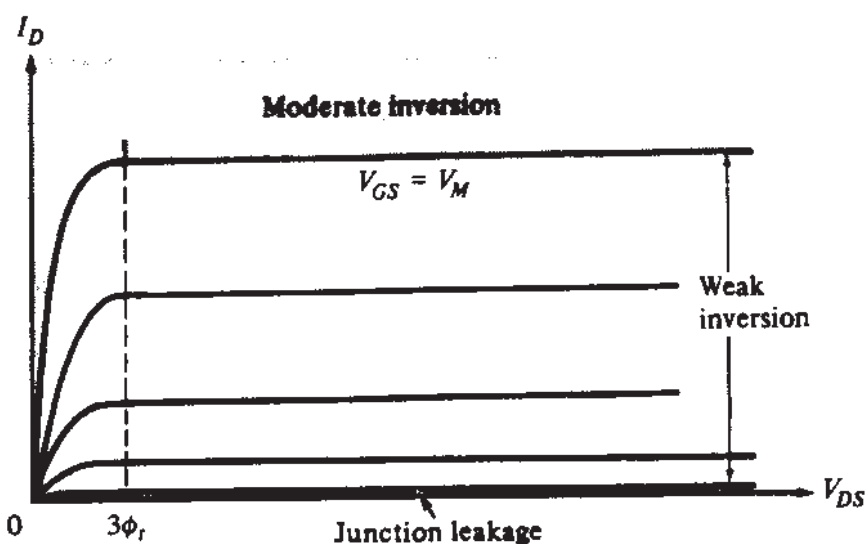


FIGURE 4.25

I_D - V_{DS} characteristics in weak inversion, with V_{GS} a parameter.

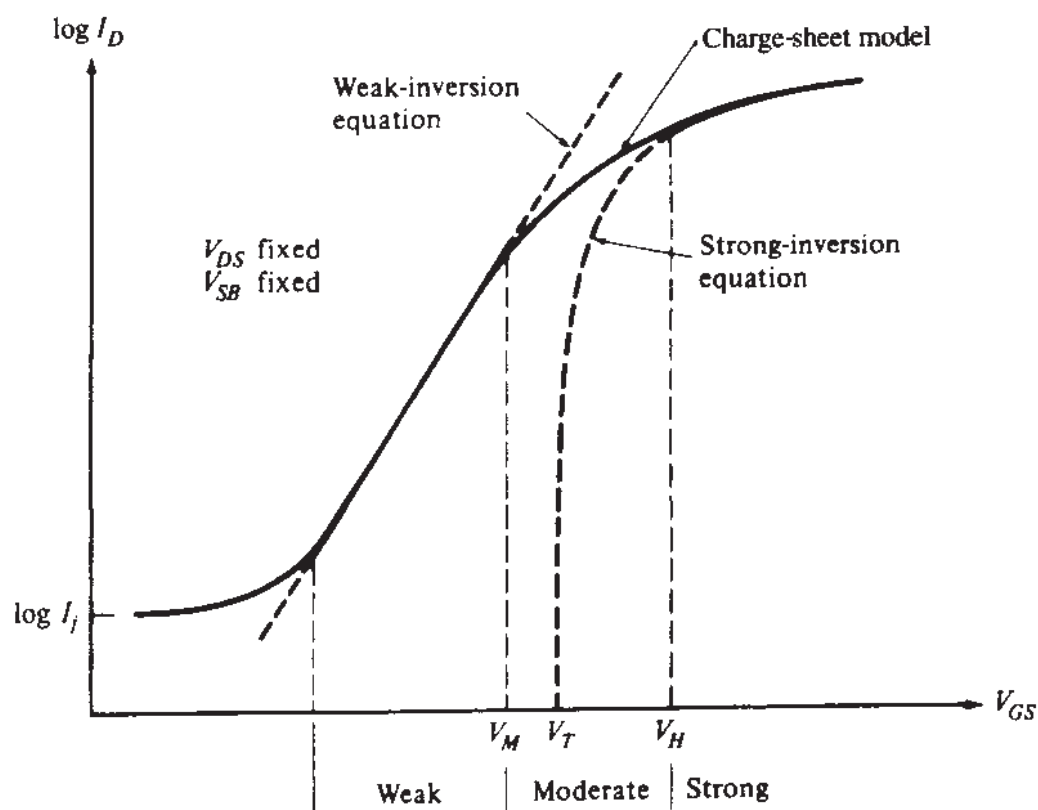


FIGURE 4.26

$\log I_D$ vs. V_{GS} for fixed V_{DS} and V_{SB} , and comparison to weak- and strong-inversion equations.

tom, I_{DS} is so small that it is masked by the leakage current I_f . At room temperatures or below, where I_f is small, the weak-inversion region can span several decades of current.

The slope of $\log I_{DS}$ vs. V_{GS} determines how well a transistor can be turned off by reducing V_{GS} , for digital applications. An often-used measure for this is the "gate

swing" S ,† defined as the amount by which V_{GS} must be reduced, in order for the weak inversion current to be reduced by 1 order of magnitude (usually expressed in "mV/decade"). It is easily seen that S is the inverse of the slope of the weak-inversion part of the plot in Fig. 4.26, i.e.,

$$S = \frac{dV_{GS}}{d(\log I_{DS})} \quad (4.6.17)$$

which, using (4.6.13), gives

$$S = 2.3n\phi_t \quad (4.6.18)$$

We emphasize that (4.6.13) is only approximate, and that I_{DS} is not exactly an exponential in weak inversion, as evident from the more accurate equations, (4.6.6)–(4.6.7). The minute deviations from exponentiality are important to keep in mind when small-signal parameters are considered (Chap. 8).

In Fig. 4.26 we also show the behavior of the strong inversion model of Sec. 4.5.3, as well as of the charge sheet model of Sec. 4.3.1.

We can now return to the question of obtaining simplified weak-inversion expressions starting from the general model of Sec. 4.3.1. Since in weak inversion the current is predominately caused by diffusion, we can work with (4.3.17) and assume $I_{DS} \approx I_{DS2}$. Unfortunately, if we use (4.6.1) and substitute $\psi_{s0} = \psi_{sa}$ and $\psi_{sL} = \psi_{sa}$ in (4.3.17), we obtain $I_{DS} = 0$! This is owing to a weakness of the *form* of this equation. In weak inversion it relies on differences of nearly equal quantities; thus, approximations such as the above are inadequate. Of course, ψ_{s0} and ψ_{sL} are *not exactly* equal; ψ_{s0} only reaches ψ_{sa} asymptotically in Fig. 4.5. To preserve appropriately the minute difference between the two quantities, we can employ (4.3.18). The procedure is outlined in Prob. 4.22 and produces exactly (4.6.6)–(4.6.7) again.

4.7 MODERATE INVERSION

As defined in Table 4.1, a MOS transistor is said to "operate in moderate inversion" if its more heavily inverted channel end is moderately inverted.

The I_{DS} - V_{DS} characteristics of the transistor in this region have a shape roughly similar to that in strong inversion, but are not described accurately by strong inversion equations since, as shown on Sec. 4.3, in moderate inversion both drift *and* diffusion contribute significantly to the value of the drain current. Convenient simplifications are not known for this region. One can use the charge sheet models of Sec. 4.3, which are valid in all regions, or use semiempirical models (see next section).

The moderate inversion region is sometimes completely ignored in the literature. This region is instead assumed to be the bottom of the strong-inversion region,

†Also called *subthreshold swing* or *subthreshold slope*. The term *subthreshold* is often used to mean weak inversion operation.

and the strong-inversion equations of Sec. 4.5 are used in it. However, not even a single point in the channel is strongly inverted in this region; hence, equations such as (4.5.1) will be in error, and so will the models based on them. In an attempt to “stretch” the validity of the strong-inversion equations, sometimes the parameters in them are made functions of the gate voltage, for example, ϕ_0 or V_T in (4.5.32).³⁴ In developing expressions for the moderate-inversion region, care should be taken that the drain current be continuous at the boundaries between weak and moderate inversion, and between moderate and strong inversion. It is also very important to make sure that the *derivatives* of the drain current with respect to any terminal voltage are *also* continuous at these boundaries; otherwise, severe errors can result in the evaluation of small signal parameters at these boundaries (Chap. 8), and convergence problems can occur during computer simulation (Chap. 10).

4.8 INTERPOLATION MODELS

Due to the difficulties mentioned above in developing moderate inversion expressions, and the need for continuity of I_{DS} and its derivatives, several semiempirical “single-piece” expressions have been proposed.^{66,68,69,80–82,89,90} These are designed to do an acceptable, or even good, job in moderate inversion, and to reduce to acceptable weak and strong inversion expressions in the corresponding regions. An example of this approach is the EKV model, which is based on the following expression:⁶⁹

$$I_{DS} = \frac{W}{L} \mu C'_{ox} (2n) \phi_t^2 \left\{ \left[\ln \left(1 + e^{(V_P - V_{SB})/(2\phi_t)} \right) \right]^2 - \left[\ln \left(1 + e^{(V_P - V_{DB})/(2\phi_t)} \right) \right]^2 \right\} \quad (4.8.1)$$

This “single-piece” model can be used in all regions, for both nonsaturation and saturation. It has a smooth behavior which approaches asymptotically the weak and strong inversion models in the respective regions. In weak inversion, both exponentials have a magnitude much smaller than 1; then by using the approximation $\ln(1+x) \approx x$ for $|x| \ll 1$, it is easily seen that the model reduces to the weak inversion symmetric model of (4.6.8), only with $2n$ in place of $(n-1)e^{(\phi_0 - 2\phi_F)/\phi_t}$ (this problem will be discussed shortly). In strong inversion and deeply in nonsaturation, both exponentials are much larger than 1. Using now the approximation $[\ln(1+e^y)]^2 \approx (\ln e^y)^2 = y^2$ for $e^y \gg 1$, we see that these exponentials produce square-law terms and the model reduces to the strong inversion expression (4.5.16). In moderate inversion no simplification is possible, and the entire expression (4.8.1) must be used.

As V_{DB} is raised, the second exponential in (4.8.1) becomes negligible; saturation is then naturally obtained, and in strong inversion the equation reduces to (4.5.19). If V_{SB} is raised instead, then it is the first exponential that becomes negligible and we enter reverse saturation.

If, in the above equation, we approximate V_p using (3.5.12), we obtain the following form^{66,68,69}

$$I_{DS} = \frac{W}{L} \mu C'_{ox} (2n) \phi_t^2 \left\{ \left[\ln \left(1 + e^{(V_{GB} - V_{T0} - nV_{SB})/(2n\phi_t)} \right) \right]^2 - \left[\ln \left(1 + e^{(V_{GB} - V_{T0} - nV_{DB})/(2n\phi_t)} \right) \right]^2 \right\} \quad (4.8.2)$$

Proceeding as above, it is easy to verify that this equation reduces to (4.5.23) (with V_{M0} replaced by V_{T0}) in strong-inversion nonsaturation, and to (4.6.9) in weak inversion, again with $(2n)$ in place of $(n-1)e^{(\phi_0 - 2\phi_F)/\phi_t}$.

Given the semiempirical nature of the above models, some error can be expected when correct physical values are used for their parameters. In such cases, a proper choice of parameter values is key to the success of the model. Consider, for example, (4.8.2), which in weak inversion reduces to (4.6.9), only with $2n$ in place of $(n-1)e^{(\phi_0 - 2\phi_F)/\phi_t}$. The latter disparity can cause an error in I_{DS} . If the value of V_{T0} in (4.6.9) is artificially increased, I_{DS} can be brought back to its correct value. Since V_{T0} appears in an exponential, it only needs to be changed by a few tens of millivolts for this to happen. One should of course be very cautious with such a practice, since such "fixes" could produce errors somewhere else. In particular, V_{T0} also affects the current in strong inversion; for example, in strong-inversion saturation (4.8.2) reduces to (4.5.24), repeated here:

$$I_{DS} = \frac{W}{L} \mu C'_{ox} \frac{1}{2n} (V_{GB} - V_{T0} - nV_{SB})^2 \quad (4.8.3)$$

Fortunately, in strong inversion the quantity in parentheses is rather large, so a small error in V_{T0} does not result in a large error in I_{DS} . In practice, during "parameter extraction," several parameter values are adjusted for a best overall fit (Chap. 10). It has been shown that for the EKV model satisfactory results can be obtained in this way.⁶⁹ The reader is cautioned, however, that using nonphysical values for model parameters can be a dangerous process in general, and that "fixing" a model problem in one aspect in this way can cause problems to appear in other aspects (e.g., fixing the accuracy in I_{DS} can worsen the accuracy in the prediction of small-signal capacitances). What is worse, it is not always clear where such problems might be caused, if one focuses only on I_{DS} accuracy during parameter extraction. Thus, a model based on physical considerations, in which physical values can be used for all its parameters, is the ideal situation. At the time of this writing, no such model exists, at least not one which expresses I_{DS} explicitly in terms of the terminal voltages in *all* regions of inversion.

One can extend the above ideas to a model that uses voltages referenced to the source, using the same type of interpolation as above,^{66,69} but modifying the expression

so that it reduces to familiar results in weak and in strong inversion.⁹¹ The resulting model is

$$I_{DS} = \frac{W}{L} \mu C'_{ox} (2n) \phi_t^2 \left\{ \left[\ln \left(1 + e^{(V_{GS} - V_T)/(2n\phi_t)} \right) \right]^2 - \left[\ln \left(1 + e^{(V_{GS} - V_T - nV_{DS})/(2n\phi_t)} \right) \right]^2 \right\} \quad (4.8.4)$$

This model makes explicit use of the threshold voltage V_T , given for each V_{SB} value by the familiar expression (4.5.33). Since V_T is explicitly used in the model, if desired, one can use for this quantity a value found from more elaborate V_T models, e.g., to take into account effects related to ion implantation (Chap. 5) or to small-channel dimensions (Chap. 6). The model is valid in all regions, including moderate inversion. In weak inversion, with both exponentials much smaller than 1 and, as before, $\ln(1+x) \approx x$ for $|x| \ll 1$, the model can be shown to reduce to (4.6.13) within a multiplicative constant (see above for a discussion of the error caused by this). Consider now strong inversion. Here the first exponential is much larger than 1. If V_{DS} is very small (i.e., we are in deep nonsaturation), the second exponential is much larger than 1, too. Using then the approximation $[\ln(1+e^y)]^2 \approx (\ln e^y)^2 = y^2$, valid for $e^y \gg 1$, and after some algebra, we obtain (4.5.37a), with n in lieu of α . If, in strong inversion, V_{DS} is raised, the second exponential becomes negligible, whereas the first exponential remains much larger than 1. We are then in saturation, and the expression simplifies to (4.5.37b), again with n in lieu of α .†

Independent of the region of inversion, as V_{DS} is raised the second exponential in (4.8.4) becomes negligible. We thus have

$$I_{DS} = \frac{W}{L} \mu C'_{ox} (2n) \phi_t^2 \left[\ln \left(1 + e^{(V_{GS} - V_T)/(2n\phi_t)} \right) \right]^2 \quad \text{saturation, all regions of inversion} \quad (4.8.5)$$

In analog circuit design, where most devices operate in saturation, this expression can be found handy for approximate calculations in any region of inversion. However, it will mostly be adequate for low-voltage work, unless n is allowed to reduce to α as strong inversion is entered.

Interpolation models of the type discussed above can be a big improvement compared to models which ignore moderate inversion altogether. They achieve a balanced compromise between accuracy and simplicity, and are sometimes preferred to complete charge sheet models (see Sec. 4.3) for reasons of computational efficiency.

†The accuracy of the model can be extended if one uses $n = 1 + \gamma/(2\sqrt{\phi + V_{SB}})$, where ϕ increases as V_{GS} is raised, so that n changes from the value in (4.6.16) to the value in (4.5.43).^{63,91}

4.9 SOURCE-REFERENCED VS. BODY-REFERENCED MODELING

In this chapter we have discussed a variety of models. Some were symmetric with respect to the substrate and gate, and the independent variables in them were V_{GB} , V_{DB} , and V_{SB} . Others were source-referenced, with V_{GS} , V_{DS} , and V_{SB} as the independent variables. Several advantages are often claimed for each approach, but it should be noted that, in principle, the two approaches are, or can be, equivalent. This is obvious, since a body-referenced model can be converted to a source-referenced one by replacing V_{GB} by $V_{GS} + V_{SB}$ and V_{DB} by $V_{DS} + V_{SB}$. Thus, all “advantages” of the body-referenced model would carry over to the new “source-referenced” model. However, in practice each approach is often combined with a different set of practices, and it is those practices that are responsible for some of the ways each approach is viewed. Let us now list the advantages of each approach. We will take this opportunity to include issues which pertain not only to modeling, but also to points of view of the MOSFET as a circuit element.

Advantages of source-referenced models:

1. Physically and intuitively, the real “driving forces” that cause the current to flow in a MOSFET are V_{GS} (to create the inversion layer) and V_{DS} (to apply a potential across that layer). They are not V_{GB} and V_{DB} per se. For a device with a desirably very small body effect, variations in V_{GB} , with V_{GS} constant, have a very small effect on I_{DS} , whereas variations of V_{GS} , with V_{GB} fixed, have a major effect. In principle, for a substrate with negligible doping concentration the B terminal would become irrelevant, and it would not make much sense to refer the voltages to it. Silicon-on-insulator MOS devices have no conducting substrate, and V_{GS} is what causes an inversion layer in them (the same is true for junction FET devices). The semiconducting substrate is a layer we would ideally do without; referring all voltages to a terminal we would rather see disappear is not easily accepted.
2. In the history of electronics, all previous major devices (vacuum tubes and bipolar transistors) were usually viewed with “input” voltages corresponding to V_{GS} (the grid-cathode voltage in tubes, and the base-emitter voltage in bipolar transistors). There was no “fourth” terminal (unless the parasitic substrate on which bipolar devices sit counts as one). Using V_{GS} as the variable for MOS devices is a natural carryover, and the intuition developed over many decades of experience with other devices can be extended to use with MOSFETs. All these types of devices can then be viewed in a unified way for the purposes of circuit design (this can be especially convenient in BiCMOS technology, where both bipolar and MOS transistors must be dealt with).
3. By referring all voltages to the source, and going through a linearization of the bulk charge, the threshold voltage occurs naturally in the device equations, as we have seen [see, for example, (4.5.37)]. The threshold voltage is one of the most extensively studied and characterized MOSFET parameters, and it is convenient to

have it as an explicit model parameter. In particular, its functional dependence on V_{SB} has been extensively studied, and the associated expressions for $V_T(V_{SB})$ can be used directly in the model equations.

4. As we have seen, the body acts as a second gate and is referred to as a "back gate" for this reason. It is thus natural to treat it similar to the way we treat the gate. In fact, there are analog circuits in which the body is actually used as a second gate (given that we cannot get rid of the substrate, we might as well use it for something). With the source as the reference, the "inputs" now are V_{GS} and V_{BS} . This can be handled naturally by a source-referenced model. For a body-referenced model, using the reference (or "ground") node B as an input is not as convenient.
5. Effects related to carrier transport from source to drain, such as velocity saturation (Chap. 6), are easier to handle in terms of V_{DS} , rather than in terms of V_{DB} and (separately) of V_{SB} .
6. Although, if a transistor is laid out symmetrically, it has symmetric characteristics, often transistors are not laid out symmetrically. For example, some gate geometries are not rectangular, and the source is made to partially surround the drain (or vice versa). In such cases, the plots of current vs. drain voltage are not exactly the same as the plots of current vs. source voltage, and the drain capacitances are certainly very different from the source capacitances. In such cases, keeping track of which terminal is the source is important, and it may be necessary to extract parameters with this in mind. Such distinction is natural to a source-referenced model. A body-referenced model can, of course, also be employed, but it will have to be made asymmetric for this purpose.
7. For work at very high frequencies, several modeling approaches have been proposed for source-referenced models (Chap. 9), whereas very little has been done in this respect for body-referenced models. Standard radio-frequency (RF) design techniques and measurements are also source-reference-based.

Advantages of body-referenced models:

1. The symmetry inherent in body-referenced models is intellectually satisfying.
2. The symmetrical nature of symmetrically laid out MOSFETs is sometimes taken advantage of in some analog circuits. A body-referenced model is well suited to use in such cases.
3. In body-referenced long-channel models, current saturation is attained at values of V_{DB} which are independent of V_{SB} , as expected from basic considerations.
4. Weak inversion is very well handled by body-referenced models. For example, the quantity ψ_{sa} in Sec. 4.6 depends only on V_{GB} . A design culture has developed using such models for circuits operating in weak inversion.³⁷
5. Some effects, like effective mobility, which are associated with transverse fields, can be very well handled in a body-referenced model.
6. A body-referenced model does not, by its very nature, exhibit discontinuities of I_{DS} or its derivatives at $V_{DS} = 0$. This property is very desirable in the implementation of models for computer simulation (Chap. 10). In contrast, source-referenced

models can present such problems, unless one is very careful. Note, though, that effects related to carrier transport from source to drain (such as velocity saturation, Chap. 6) are often handled through V_{DS} even in body-referenced models; in such formulations, these models can also present discontinuities at $V_{DS} = 0$.

Despite the above differences, and as already indicated in the beginning of this subsection, source-referenced and body-referenced models can in principle be equivalent. Either approach can, in principle, do an equally good job, and the user of a model implemented in a circuit simulator may in fact not even need to know whether the internal structure of the model is source-referenced or body-referenced. It is the practices usually associated with each type, as well as differences in circuit design cultures, that are responsible for preferring one or the other type of model in several instances. We find it appropriate to discuss both types of model in this book.

4.10 EFFECTIVE MOBILITY

In Sec. 1.3.2, we briefly looked at the mobility of electrons in the bulk. We mentioned that the value of this quantity is determined by several scattering mechanisms, through which the electrons exchange momentum and kinetic energy with their environment. One such mechanism is due to the energy of lattice vibrations, which is modeled by using energy quanta called *phonons*. The electron-phonon interaction is referred to as *phonon scattering*. Another mechanism is related to the electrically charged ionized impurity atoms, and is called *Coulomb scattering*.

In a MOS transistor, electrons in the inversion layer flow near the semiconductor-oxide interface (the “surface” of the semiconductor). The electric field component perpendicular to the direction of current flow (referred to as the *transverse*, or *normal*, component) tends to accelerate the inversion layer electrons toward the surface and subjects them to additional scattering.^{17,92} Now there is Coulomb scattering not only due to ionized impurity atoms, but also due to interface trapped charges, and to charges trapped within the oxide. It is interesting to note that the inversion layer charge itself, if it is significant, tends to partly “screen” itself from the effects of Coulomb scattering. Additional scattering occurs due to surface roughness. All these tend to lower the mobility of electrons in the inversion layer (called *surface mobility*) to values smaller than the bulk mobility considered in Sec. 1.3.2.†

Surface mobility has been the subject of many studies.^{17,92–139,55,62} In most of these, surface mobility is studied as a function of an average transverse field; while this quantity can be defined in more than one way, we will assume for now that it is the mean between the values of the transverse field at the top and at the bottom of the inversion layer. A typical behavior of surface mobility vs. average transverse field is as shown in Fig. 4.27. Qualitatively, this behavior can be described by taking into

†In short-channel devices, things get further complicated due to the large field along the surface (the parallel, or longitudinal field); this will not concern us until we get to Chap. 6.

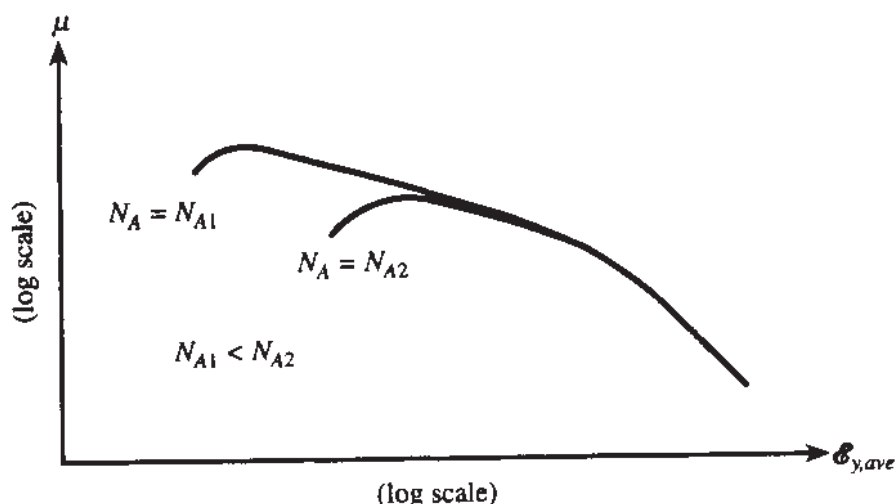


FIGURE 4.27
Surface mobility vs. average transverse field.

account the fact that, while all the above mechanisms may be present simultaneously, and while each tends to limit the surface mobility value, different mechanisms tend to dominate the behavior in different ranges of field values. At low fields, Coulomb scattering due to impurity atoms and oxide charges tends to be dominant in limiting mobility.^{121,131-133} As the field is increased, the inversion layer charge becomes significant and tends to screen the impurity and oxide charges, thereby lessening their effect; this is why the mobility shows an initial increase in the figure. This trend does not continue as the field is increased further, since phonon scattering becomes important and mobility begins to drop. Finally, at still higher fields, the strong "pull" of electrons toward the surface makes surface roughness the dominant scattering mechanism, and mobility begins to drop faster, as seen toward the right.

Increasing the doping concentration tends to shift the point where Coulomb scattering becomes nondominant to higher field values. This is because there are now more ionized impurity atoms, which makes Coulomb scattering more prominent; also, higher substrate dopings mean that a larger substrate charge must be depleted [see (3.2.5a)] before an inversion layer can be formed, which means that the inversion layer charge will not become significant (so that it can screen the effects of Coulomb scattering), until higher field values.¹²⁷ Thus the peak of the mobility curve shifts to the right if doping is increased, as shown. We should note here that the literature is not consistent as to how serious the mobility drop is at low electric fields, and in fact some or all of this drop, as observed by some authors, has been claimed by others^{108,124} to be the result of experimental error, or of errors in data interpretation.

It has been reported that to the right of their peak, the curves for different doping concentrations and different oxide thicknesses tend to merge and to follow a universal curve,^{98,100,101,103,109,127,139} although surface orientation and surface preparation can make a difference.¹²⁸ A simple approximate equation for this "universal" behavior will be presented later in this section.

We will now study ways to incorporate the above behavior in MOS transistor models. Recall that, in steady state, the drain current is equal to the current at any point in the channel; including drift and diffusion components we have (4.3.5), repeated below:

$$I_{DS} = \mu W (-Q'_l) \frac{d\psi_s}{dx} + \mu W \phi_t \frac{dQ'_l}{dx} \quad (4.10.1)$$

In Sec. 4.3, we integrated the above equation from $x = 0$ to $x = L$, obtaining

$$I_{DS} = \frac{W}{L} \left[\int_{\psi_{s0}}^{\psi_{sL}} \mu (-Q'_l) d\psi_s + \phi_t \int_{Q'_{l0}}^{Q'_{lL}} \mu dQ'_l \right] \quad (4.10.2)$$

We then made the assumption that μ was constant along the channel; thus μ could be moved outside the integral, and we obtained

$$I_{DS} = \frac{W}{L} \mu \left[\int_{\psi_{s0}}^{\psi_{sL}} (-Q'_l) d\psi_s + \phi_t (Q'_{lL} - Q'_{l0}) \right] \quad (4.10.3)$$

However, we now know that μ depends on the transverse field. Since this field, in general, varies along the channel, μ will also vary. If its variation is not negligible, one *cannot* move μ outside the integral. One can then attempt to determine μ as a function of the variables of integration in (4.10.2), and try to include this function in the integration. This is very difficult analytically.²⁴ Another approach taken is to define an *effective* mobility μ_{eff} such that, when used in the expression

$$I_{DS} = \frac{W}{L} \mu_{\text{eff}} \left[\int_{\psi_{s0}}^{\psi_{sL}} (-Q'_l) d\psi_s + \phi_t (Q'_{lL} - Q'_{l0}) \right] \quad (4.10.4)$$

the current is predicted correctly.[†] Since the electric field at each point in the channel and, thus, μ in (4.10.2) depend on the terminal voltages, one should expect that to make (4.10.4) give the same result as (4.10.2), μ_{eff} should be made a function of the terminal voltages. This function is often determined empirically, but sometimes a more analytical approach can be used, as will be shown. Comparing (4.10.4) to (4.10.3) we see that, when the effective-mobility approach is taken, I_{DS} will be given by the same expressions found in Sec. 4.3, if in them μ is replaced by μ_{eff} :

| | | | |
|--|---|---|----------|
| $I_{DS, \text{ including mobility}}$ dependence on normal field | $= I_{DS, \text{ assuming constant}}$ mobility | $\left \mu = \mu_{\text{eff}} \right.$ | (4.10.5) |
|--|---|---|----------|

[†]A formal definition of μ_{eff} can be obtained by equating the right-hand sides of (4.10.4) and (4.10.2), and solving for μ_{eff} . However, this is not useful for evaluating μ_{eff} since, if one can evaluate the integrals in (4.10.2) with adequate simplicity, then μ_{eff} is not needed; one would, in this case, have found I_{DS} directly from (4.10.2).

We now describe a way to determine μ_{eff} in principle; this is an extension⁶⁵ of classical techniques to the charge sheet model. If both sides of (4.10.1) are divided by μ , and then integration is performed from $x = 0$ to $x = L$, we obtain

$$I_{DS} \int_0^L \frac{dx}{\mu} = W \left[\int_{\psi_{s0}}^{\psi_{sL}} (-Q'_I) d\psi_s + \phi_t (Q'_{IL} - Q'_{I0}) \right] \quad (4.10.6)$$

Solving this for I_{DS} , we find an expression which is exactly of the form of (4.10.4), with

$$\mu_{\text{eff}} = \frac{1}{\frac{1}{L} \int_0^L \frac{dx}{\mu}} \quad (4.10.7)$$

As a check we note that, if μ is assumed constant, the above equation gives $\mu_{\text{eff}} = \mu$, as expected.

For the real case where μ depends on \mathcal{E}_y , one can in principle express \mathcal{E}_y in terms of the surface potential, and relate the latter to the position x through (4.3.22). Whether this will give manageable results from (4.10.7) depends on the form assumed for $\mu(\mathcal{E}_y)$, and on the assumptions made along the way.

As an example, we now show a popular way of doing this in strong inversion. In this region, mobility is assumed to follow the "universal" behavior shown to the right of the peaks in Fig. 4.27, and is written as a function of:^{24,94}

$$\mathcal{E}_{y,\text{ave}} = \frac{\mathcal{E}_{ys} + \mathcal{E}_{yb}}{2} \quad (4.10.8)$$

where \mathcal{E}_{ys} is the value of the normal field at the surface and \mathcal{E}_{yb} is its value "just below" the inversion layer; these values will be discussed shortly.

Experimental data can be approximately modeled by the following relation:⁹⁴

$$\mu = \frac{\mu_0}{1 + a_\theta \mathcal{E}_{y,\text{ave}}} \quad (4.10.9)$$

where μ_0 and a_θ depend on temperature and can be considered fitting parameters.[†] For practical conditions, μ_0 is roughly half of the bulk mobility (as evaluated for lightly doped substrates) and a_θ is of the order of 10^{-6} cm/V at room temperature.

[†]It is interesting to note that the form of (4.10.9) is not totally empirical. It can be made plausible by starting from the so-called "Mathiessen rule" for combining mobility contributions, which gives $1/\mu = 1/\mu_1 + 1/\mu_2$, with μ_1 the mobility due to bulk scattering and μ_2 the mobility due to surface scattering; contributions to bulk mobility are combined in a similar manner.^{16,58} If we assume that μ_2 is inversely proportional to $\mathcal{E}_{y,\text{ave}}$ (which, in turn, can be made plausible under certain physical assumptions),⁹³ it is easily seen that the form of (4.10.9) results. However, the values of the parameters in that equation are difficult to justify rigorously. A semiempirical discussion is given elsewhere.⁹³ Note that (4.10.9) does not model the post-peak behavior at very high fields.¹²⁷ A better fit to experimental data is claimed to be obtained by $\mu = \mu_0 / (1 + a_1 \mathcal{E}_{y,\text{ave}}^{b_1} + a_2 \mathcal{E}_{y,\text{ave}}^{b_2})$, where a_1 and a_2 are fitting parameters, b_1 is close to 0.3, and b_2 is approximately 2 for electrons and 1 for holes.^{122,126} We note, however, that the stronger degradation of mobility at very high fields is not universally accepted.¹⁰¹

The electric field at the surface can easily be related to the total charge per unit area below the surface from basic electrostatics (Appendix B)†:

$$\mathcal{E}_{ys} = -\frac{Q'_I + Q'_B}{\epsilon_s} \quad (4.10.10)$$

where Q'_I and Q'_B are the inversion layer and depletion region charges per unit area, respectively, and ϵ_s is the permittivity of the semiconductor. Similarly, the field just below the inversion layer can be determined by assuming a very thin inversion layer, so that the total charge per unit below the inversion layer is practically all of Q'_B :

$$\mathcal{E}_{yb} = -\frac{Q'_B}{\epsilon_s} \quad (4.10.11)$$

Using the above equations in (4.10.8), we obtain‡

$$\mathcal{E}_{y,ave} = -\frac{Q'_B + 0.5Q'_I}{\epsilon_s} \quad (4.10.12)$$

and thus (4.10.9) becomes

$$\mu = \frac{\mu_0}{1 - (\alpha_\theta/\epsilon_s)(Q'_B + 0.5Q'_I)} \quad (4.10.13)$$

Substituting this in (4.10.7) we have:

$$\mu_{eff} = \frac{\mu_0}{\frac{1}{L} \int_0^L [1 - (\alpha_\theta/\epsilon_s)(Q'_B + 0.5Q'_I)] dx} \quad (4.10.14)$$

Unfortunately, the calculation of this is not easy since Q'_I and Q'_B depend on x through the channel-substrate effective reverse bias $V_{CB}(x)$, which is, in general, given by a complicated expression; see, for example, (4.5.49). It is here that a simplification is made: For the purposes of evaluating (4.10.14), it is assumed¹³⁹ that V_{CB} varies approximately linearly with x . (As seen from Fig. 4.22 this assumption is

†Due to the gradual channel approximation (Sec. 4.1), the electric field component parallel to the direction of current flow is assumed negligible compared to the transverse component. Hence, a one-dimensional analysis can be used.

‡The appropriateness of the above simple definition of $\mathcal{E}_{y,ave}$ in (4.10.8) and the resulting expression for it in (4.10.12) has been questioned.¹²⁵ For p -channel devices, it has been found that the "universality" of the mobility-field behavior is wider if the factor of 0.5 in (4.10.12) is replaced by a smaller factor⁵⁷ in the range of 0.25 to 0.38. The limits of mobility-field "universality" have been discussed.^{118,127,128,130} We note that such universality is less valid for heavier substrate doping.¹²⁷

mostly satisfactory for low values of V_{DS} .) Then we can write $dV_{CB}/dx \approx (V_{DB} - V_{SB})/L$. Using this to perform a change of variables in the integral of (4.10.14), we obtain

$$\mu_{\text{eff}} \approx \frac{\mu_0}{\left[1/(V_{DB} - V_{SB})\right] \int_{V_{SB}}^{V_{DB}} \{1 - (\alpha_\theta/\epsilon_s)(Q'_B + 0.5Q'_I)\} dV_{CB}} \quad (4.10.15)$$

The integration can be completed by using Q'_I and Q'_B from (4.5.10a) and (4.5.9) for the complete strong-inversion model;¹³⁹ for the simplified strong-inversion model, the corresponding equations to use are (4.5.30) and (4.5.29). After some algebra, we obtain

$$\boxed{\mu_{\text{eff}} = \frac{\mu_0}{1 + \theta f_\mu}} \quad (4.10.16)$$

where

$$\theta = \frac{\alpha_\theta}{2\epsilon_s} C'_{\text{ox}} \quad (4.10.17)$$

and f_μ is given by the following expressions.

*For the complete strong-inversion model:*¹³⁹

$$f_\mu = (V_{GB} - V_{FB} - \phi_0) - \frac{1}{2}(V_{DB} + V_{SB}) + \frac{2}{3}\gamma \frac{(\phi_0 + V_{DB})^{3/2} - (\phi_0 + V_{SB})^{3/2}}{V_{DB} - V_{SB}} \quad (4.10.18)$$

For the simplified strong-inversion model:

$$f_\mu = V_{GS} - V_{FB} - \phi_0 + \gamma\sqrt{\phi_0 + V_{SB}} - \left(1 - \frac{\alpha}{2}\right)V_{DS} \quad (4.10.19a)$$

$$= V_{GS} - V_T + 2\gamma\sqrt{\phi_0 + V_{SB}} - \left(1 - \frac{\alpha}{2}\right)V_{DS} \quad (4.10.19b)$$

where V_T is given by (4.5.33).

As seen, the effect of the gate voltage is dominant. For this reason, it is sometimes said that μ_{eff} "depends on the gate field." It is more correct, though, to say that

†The fraction in this equation becomes 0/0 for $V_{DB} = V_{SB}$. To avoid this problem, one can factor out $(\phi_0 + V_{DB})^{1/2} - (\phi_0 + V_{SB})^{1/2}$ from both numerator and denominator, and replace the fraction by⁵⁹

$$\frac{(\phi_0 + V_{DB}) + (\phi_0 + V_{DB})^{1/2}(\phi_0 + V_{SB})^{1/2} + (\phi_0 + V_{SB})}{(\phi_0 + V_{DB})^{1/2} + (\phi_0 + V_{SB})^{1/2}}$$

μ_{eff} depends on the transverse field, which, in turn, depends on all terminal voltages, as has been seen. We should note here that in the denominator of the expression for μ_{eff} another term proportional to V_{DS} is often included to model "velocity saturation" effects. This is not related to the V_{DS} terms in the above equations, and will be considered in detail in Chap. 6.

Using μ_{eff} from (4.10.16) and (4.10.18) in lieu of μ in the accurate strong-inversion expression (4.5.2) has been shown to result in an accurate model with excellent agreement to experiment.¹³⁹ One can also use (4.10.16) with (4.10.19) in conjunction with (4.5.31). In either case, of course, one must use the approach indicated in Sec. 4.5 to extend the characteristics into the saturation region. In the latter case, for example, one needs to determine the value of V_{DS} at which saturation is obtained. This value, denoted by V'_{DS} , is the value of V_{DS} at which $dI_{DSN}/dV_{DS} = 0$. Note that, because of the presence of V_{DS} in the expression for μ_{eff} , a new differentiation is required to obtain the value of V'_{DS} . The V'_{DS} expressions of Sec. 4.5 will not be valid here, since they were obtained by assuming a constant mobility. This results in additional complications which are often avoided by neglecting the dependence of μ_{eff} on V_{DS} . For example, consider (4.10.19b). In that equation the last term is usually dropped.[†] In addition, the term dependent on V_{SB} is sometimes replaced by a term linear in V_{SB} . Thus the following form has been suggested for (4.10.16)^{25,85‡}

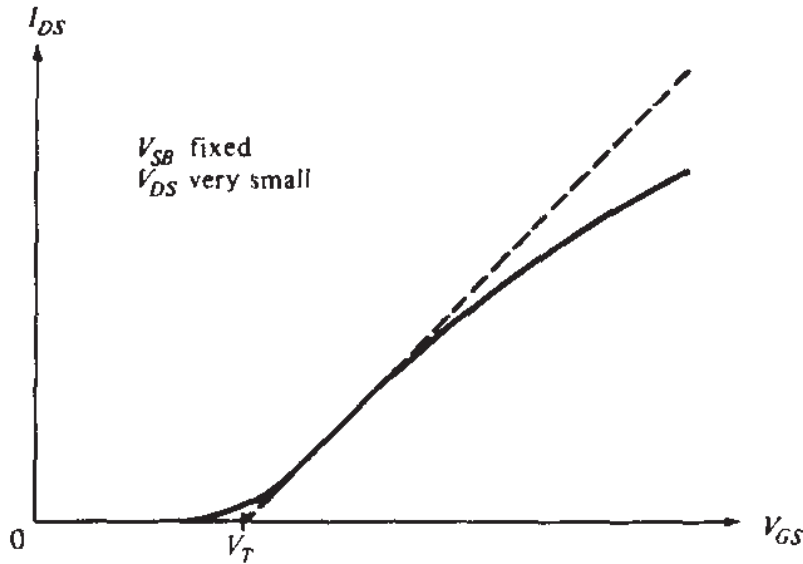
$$\mu_{\text{eff}} = \frac{\mu_0}{1 + \theta(V_{GS} - V_T) + \theta_B V_{SB}} \quad (4.10.20)$$

By now several approximations have been made, and one should expect that the values of μ_0 , θ , and θ_B used in the above equation may have to be chosen by fitting to measurements, to minimize the error. A typical value for μ_0 is $600 \text{ cm}^2/(\text{V} \cdot \text{ns})$ for n -channel devices at room temperature. For p -channel devices, μ_0 is typically smaller by a factor of 3. The parameter θ is of the form $\beta_\theta/t_{\text{ox}}$ with t_{ox} the oxide thickness, and β_θ typically 5 to $20 \text{ \AA} \cdot \text{V}^{-1}$.§ The value of θ_B is usually small (for example, a few hundredths of 1 V^{-1}), and often the V_{SB} -dependent term in (4.10.20) is omitted altogether. This can cause problems, as then increasing V_{SB} would imply an increase in μ_{eff} , since V_T would be increasing owing to the body effect. However, μ_{eff} should actually *decrease* with increasing V_{SB} . This is most easily seen from (4.10.16) with (4.10.19a) and is also expected intuitively. Let us refer all voltages to the source. Then increasing V_{SB} means making the voltage at the substrate terminal more negative. This would tend to increase the normal field and "push" the electrons more

[†]Often the effect of this term is absorbed into a similar term that results from "velocity saturation" effects; these are considered in Chap. 6.

[‡]A term containing V_{GS}^2 is sometimes added^{114,118} to the denominator of μ_{eff} , to help model the strong decrease of mobility due to surface roughness at high fields, such as those encountered in transistors with very thin oxides, although the presence of this effect is not universally accepted.

[§]The value of θ is sometimes adjusted to help model an effect quite unrelated to the field dependence of mobility, i.e., the effective drain-source voltage reduction due to the voltage drop across the series resistance of the source and drain regions. This effect is discussed in Sec. 6.8.

**FIGURE 4.28**

I_{DS} vs. V_{GS} for fixed V_{SB} and fixed, very small V_{DS} for constant effective mobility and neglecting moderate and weak inversion (broken line), and V_{GS} -dependent effective mobility and including moderate and weak inversion (solid line).

toward the surface, which is the same effect an increase in the gate voltage would have; thus μ_{eff} should be expected to decrease. This is verified experimentally¹³⁹ and is found to be more evident in p -channel devices (Sec. 4.1).⁸⁵ Thus omitting the V_{SB} -dependent term in (4.10.20) can have a serious effect, especially in the small-signal modeling of analog circuits (Chaps. 8 and 9). If, in addition to omitting the V_{SB} -dependent term in (4.10.20) V_T is replaced by V_{T0} [as given by (4.5.33)],⁸⁶ μ_{eff} appears independent of V_{SB} .

The effect of mobility dependence on V_{GS} is shown in Fig. 4.28. I_{DS} is plotted vs. V_{GS} for a fixed V_{SB} and very small V_{DS} . Then, as seen from (4.5.37), we will have $I_{DS} \approx (W/L)\mu C'_{\text{ox}} V_{DS}(V_{GS} - V_T)$. If μ is constant, a straight line is obtained in strong inversion, as shown by the broken line (the bottom, curved part of the solid curve is caused by moderate inversion). If μ is replaced by μ_{eff} , which varies with V_{GS} as suggested above, the plot becomes as shown by the solid line. In some devices this effect is so strong that one never really gets to see a straight-line part in the characteristic. Note that very small V_{DS} is not a necessary condition for obtaining a straight line in *nonsaturation*. However, such plots are usually obtained for very small V_{DS} since, then, nonsaturation is observed even at small V_{GS} values, at which one can hope to see the initially straight part of the curve where mobility degradation is not strong. Also, for very small V_{DS} , the intercept of an extension of the straight-line part with the horizontal axis is approximately V_T . This is a convenient method for determining the threshold voltage experimentally.

The parameter μ_{eff} is often simply referred to as the *mobility* instead of the more complete *effective mobility*, and is denoted by μ for simplicity. For convenience, we will adopt this notation in the rest of this book. No confusion should arise because it should be understood that, every time μ is encountered in expressions that give the drain current as a function of the terminal voltages, it will be the *effective mobility*.

As already pointed out, the results obtained in the above example are valid in the strong-inversion region, where the normal electric field is not very small and (4.10.9) is valid. At very low fields (e.g., corresponding to moderate or weak inversion) this equation is no longer claimed to be valid.¹³⁴ It has been suggested that the electron density in the inversion layer at such fields is no longer large enough to "screen" the ionized acceptor atoms, or that it can show significant fluctuations due to the presence of localized charges at the oxide-semiconductor interface.^{96,93} Devices have been reported to exhibit a drop in effective mobility when operating in weak inversion,^{92,93,95,96,129} although not all results reported in the literature show such behavior.^{108,134} In contrast to strong inversion, not much work has been done toward developing simple expressions for the effective mobility in weak inversion.

For readers who plan to search the literature on the important topic of effective mobility, we would like to offer some warnings:

1. What is often called "effective mobility" in the literature is not meant in the general sense we have used it above, which allowed us to write (4.10.5) for all terminal voltage combinations, including the effect of large V_{DS} on the transverse electric field. Rather, what is often meant by this term is only the mobility at a point in the channel, and it is measured indirectly, by applying a very small V_{DS} (e.g., 10 mV) on a transistor, in an attempt to make the channel uniform.
2. Some of the data on effective mobility vs. field behavior published in the literature are distorted due to experimental errors.
3. Some studies are inconclusive, and different studies contradict each other; we gave some indication of this wherever possible.
4. Many of the detailed results on effective mobility vs. field are not usable in developing a device I - V model of manageable simplicity. Thus, one often ends up using simple approximations such as the ones we have presented. One should not be surprised, therefore, if these do not provide very accurate results, or if the parameters used must be given nonphysical values in order to help make up for the lack of accuracy.

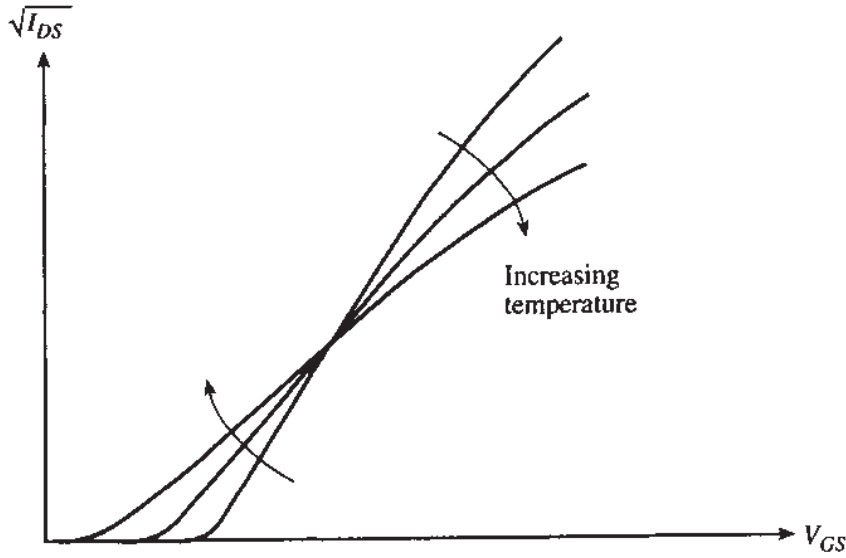
4.11 TEMPERATURE EFFECTS

MOS transistor characteristics are strongly temperature-dependent.^{18,83,85,57,58,118,140-146} One of the main parameters responsible for this is the effective mobility, which is known to decrease with temperature. An often used approximation is

$$\mu(T) = \mu(T_r) \left(\frac{T}{T_r} \right)^{-k_3} \quad (4.11.1)$$

where T is absolute temperature, T_r is room absolute temperature, and k_3 is a constant, with various values used for it^{85,57,58,94,118} between 1.2 and 2.0.

Other temperature-dependent parameters are ϕ_0 and V_{FB} [the latter is temperature-dependent through ϕ_{MS} in (2.2.6), assuming Q'_o is fixed]. These effects are manifested in

**FIGURE 4.29**

$\sqrt{I_{DS}}$ in saturation vs. V_{GS} for three different temperatures.

the value of V_T in (4.5.32), which is found to exhibit an almost straight-line decrease with temperature^{18,57,118,140,141,143,144} and can be approximated by

$$V_T(T) = V_T(T_r) - k_4(T - T_r) \quad (4.11.2)$$

where k_4 is usually between 0.5 mV/K and 3 mV/K, with larger values in this range corresponding to heavier doped substrates, thicker oxides, and smaller values of V_{SB} .

As an example of the effect of temperature on transistor characteristics, consider a device operating in the saturation region. From (4.5.37) we have

$$\sqrt{I_{DS}} = \sqrt{\mu(T)} \sqrt{\frac{1}{2} \frac{W}{L} \frac{C'_{ox}}{\alpha}} [V_{GS} - V_T(T)] \quad (4.11.3)$$

Thus, a temperature increase should tend to increase the drain current through $V_{GS} - V_T(T)$, and to decrease it through $\mu(T)$. A set of $\sqrt{I_{DS}}$ vs. V_{GS} curves in saturation is shown in Fig. 4.29. At high currents, the decrease of $\mu(T)$ with temperature dominates; the opposite is true at low currents. In certain cases a value of V_{GS} can be found at which the current becomes practically temperature-independent over a large temperature range.^{18,140,143} This effect is evident in the figure. The bottom, curved part of the curves is due to moderate and weak inversion; the curving of the upper part is due to the dependence of effective mobility on V_{GS} (Sec. 4.10).

As can be deduced from the figure, in weak inversion and for given V_{GS} the drain current increases with temperature. Typical plots of $\log I_D$ vs. V_{GS} are shown for various temperatures in Fig. 4.30. As seen, increasing temperature decreases the slope of the curves. Also, the junction leakage (the effect of which is shown by broken curves) drastically increases with temperature and masks the weak-inversion current, thus diminishing the range of currents over which near-exponential behavior is observed.

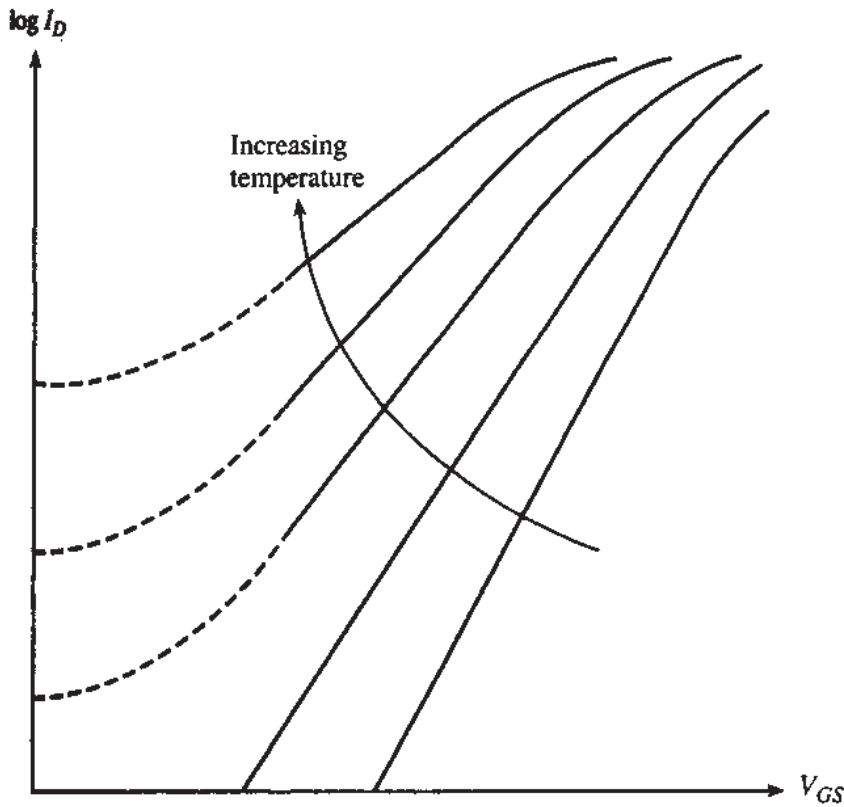


FIGURE 4.30

$\log I_D$ vs. V_{GS} at low currents for various temperatures. The broken part indicates the effect of leakage.

The reader is referred to the literature for considerations applying to operation at very low^{145,146} and very high¹⁴³ temperatures.

4.12 BREAKDOWN

The value of the various voltages that can be applied to a MOS transistor should be limited to avoid several forms of breakdown.^{18,83} One such form is *junction breakdown*. The junctions formed by the substrate and the drain or source regions will conduct a large current if the reverse bias applied to them exceeds a certain value (because the field in the junctions near the surface is influenced by the presence of the gate, the above value depends on the gate potential and can be different from that predicted by common *pn* junction theory). Junction breakdown will occur even with the device off. When the device is on, carriers moving fast in the channel can impact on silicon atoms and ionize them, producing electron-hole pairs; this is referred to as *impact ionization*. The newly generated pairs can gain enough energy to impact on silicon atoms and produce more electron-hole pairs, etc. This is called the *avalanche* effect and is more pronounced in the pinchoff region near the drain where fields can be high. Currents larger than those predicted by common device models will then flow, and the phenomenon is referred to as *channel breakdown*. More details on this phenomenon are given in Chap. 6.

The above forms of breakdown are nondestructive. Once the large voltages producing them are removed, the device will function properly as long as no permanent

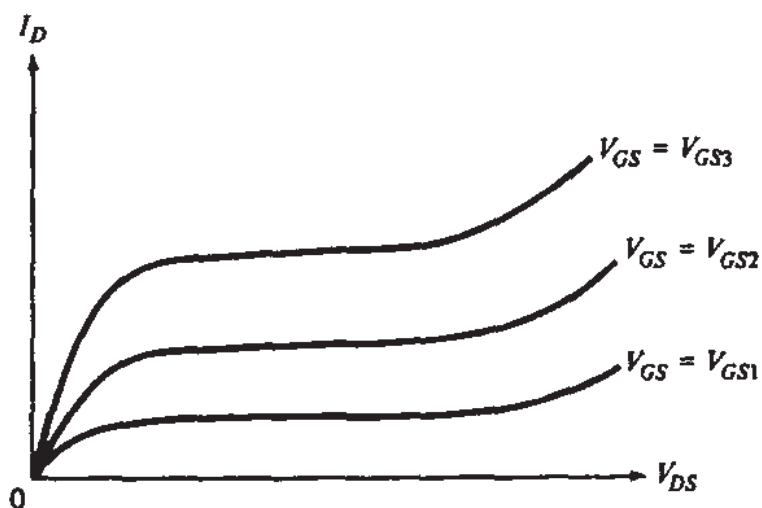


FIGURE 4.31
Effect of breakdown on I_D - V_{DS} characteristics.

damage caused by overheating has occurred. The effect of such breakdown on device characteristics is shown in Fig. 4.31.†

A *destructive breakdown* mechanism is *oxide breakdown*. It occurs when the electric field in the gate insulator exceeds a certain value [about 7×10^6 V/cm (0.07 V/Å) in silicon dioxide]. The result is a permanent short circuit through the insulator. Static charge, such as that transferred to gates by handling devices with bare hands, is known to cause oxide breakdown. For this reason, protective devices are used at those input terminals of an MOS integrated circuit that are connected to transistor gates.

4.13 THE *p*-CHANNEL MOS TRANSISTOR

If the substrate is made of *n*-type material, and the source/drain regions of *p*⁺-type material, we have what is known as the *p-channel* MOS transistor or PMOS transistor. Figure 4.32 shows such a device. An example of *p-channel* transistor characteristics is shown in Fig. 4.33.

The operation of the *p-channel* transistor is the “dual” of *n-channel* operation. The role of electrons is played by holes, and the role of ionized acceptor atoms is played by ionized donor atoms. Statements made about *n-channel* devices can be adapted to the case of *p-channel* devices with simple modifications. For example, in Fig. 4.32, the more *negative* the gate-source voltage the heavier the concentration of *holes* near the surface. The more *negative* the drain-source voltage the heavier the flow of *holes* from source to drain; hence, the more *negative* the drain current (assuming the reference direction is chosen as before, i.e., from the drain through the channel to the source). The more *negative* the source-substrate bias the larger the number of *donor* atoms that are depleted. The body effect coefficient γ is given by

$$\gamma = \frac{\sqrt{2q\epsilon_s N_D}}{C'_{ox}} \quad (4.13.1)$$

where N_D is the concentration of donor atoms in the substrate.

†Our discussion here is limited to long-channel devices. An additional form of breakdown, called *punchthrough*, is encountered in short-channel devices; this effect is considered in Sec. 6.4.

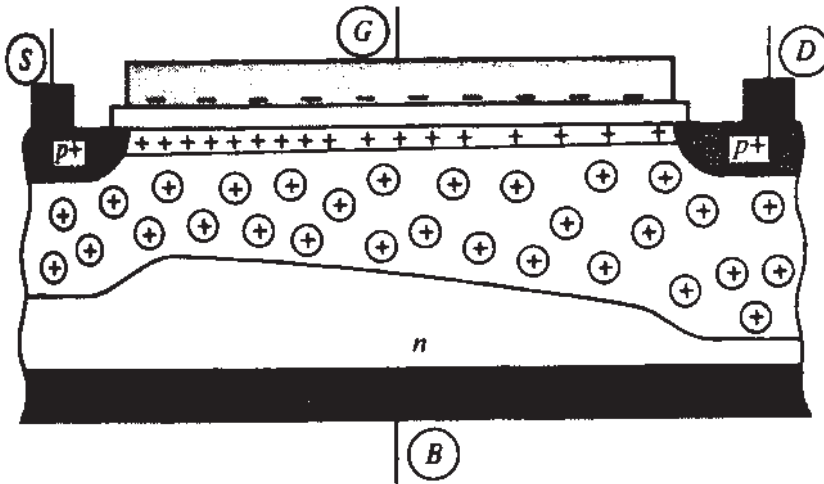


FIGURE 4.32
A *p*-channel MOS transistor.

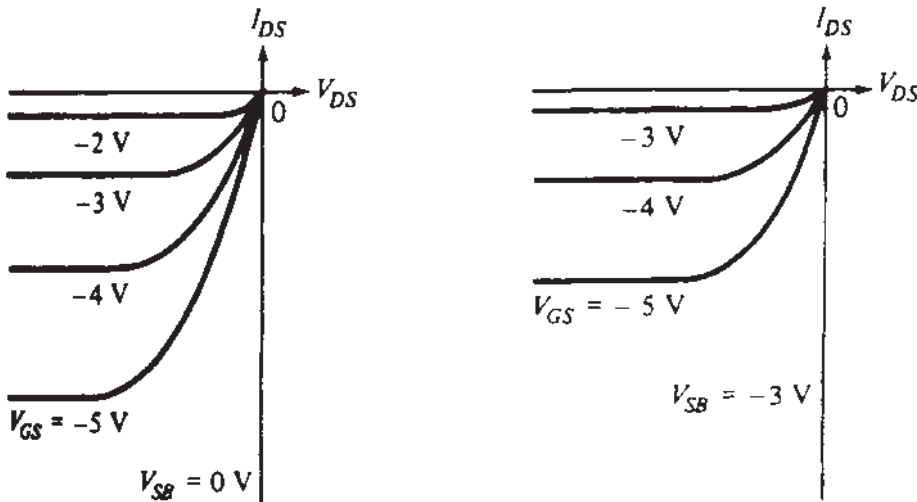


FIGURE 4.33
 I_{DS} - V_{DS} characteristics for a *p*-channel MOS transistor for two different V_{SB} values.

In describing *p*-channel devices, there are some rather obvious sign changes in the model equations; for example, instead of (4.5.31) we will have

$$I_{DSN} = -\frac{W}{L} \mu C'_{ox} \left[(V_{GS} - V_T) V_{DS} - \frac{\alpha}{2} V_{DS}^2 \right] \quad (4.13.2)$$

where instead of (4.5.33a) we will have

$$V_T(V_{SB}) = V_{T0} - \gamma (\sqrt{-V_{SB} - \phi_0} - \sqrt{-\phi_0}) \quad (4.13.3)$$

where V_{SB} and ϕ_0 are negative and, instead of (4.5.33b), we will have

$$V_{T0} = V_{FB} + \phi_0 - \gamma \sqrt{-\phi_0} \quad (4.13.4)$$

The value of the effective mobility for p -channel devices at low gate voltages is smaller than that in n -channel devices by a factor of 2 to 4; a typical value is $250 \text{ cm}^2/(\text{V}\cdot\text{s})$.

4.14 ENHANCEMENT-MODE AND DEPLETION-MODE TRANSISTORS

Consider the simple model of (4.5.37). Depending on the sign of V_{T0} , MOS transistors are separated into two categories. The n -channel transistors with positive V_{T0} are called *enhancement-mode* (or “normally off”) devices, whereas n -channel transistors with negative V_{T0} are said to be *depletion-mode* (or “normally on”) devices. These names originated in the days when weak inversion was neglected, and the n -channel transistor was viewed as being on for $V_{GS} > V_{T0}$ and off for $V_{GS} < V_{T0}$ (with $V_{SB} = 0$ assumed). Thus, if V_{T0} is positive, an n -channel device is assumed off at $V_{GS} = 0$, and it takes a positive V_{GS} to “enhance” the channel and turn the device on. If V_{T0} is negative, the n -channel device is already on with $V_{GS} = 0$, and it takes a negative V_{GS} to “deplete” the channel and turn the device off. For p -channel devices, a negative V_{T0} corresponds to the enhancement mode and a positive V_{T0} to the depletion mode. The I_D - V_{GS} characteristics for the four types of devices for a very small $|V_{DS}|$ are shown in Fig. 4.34.

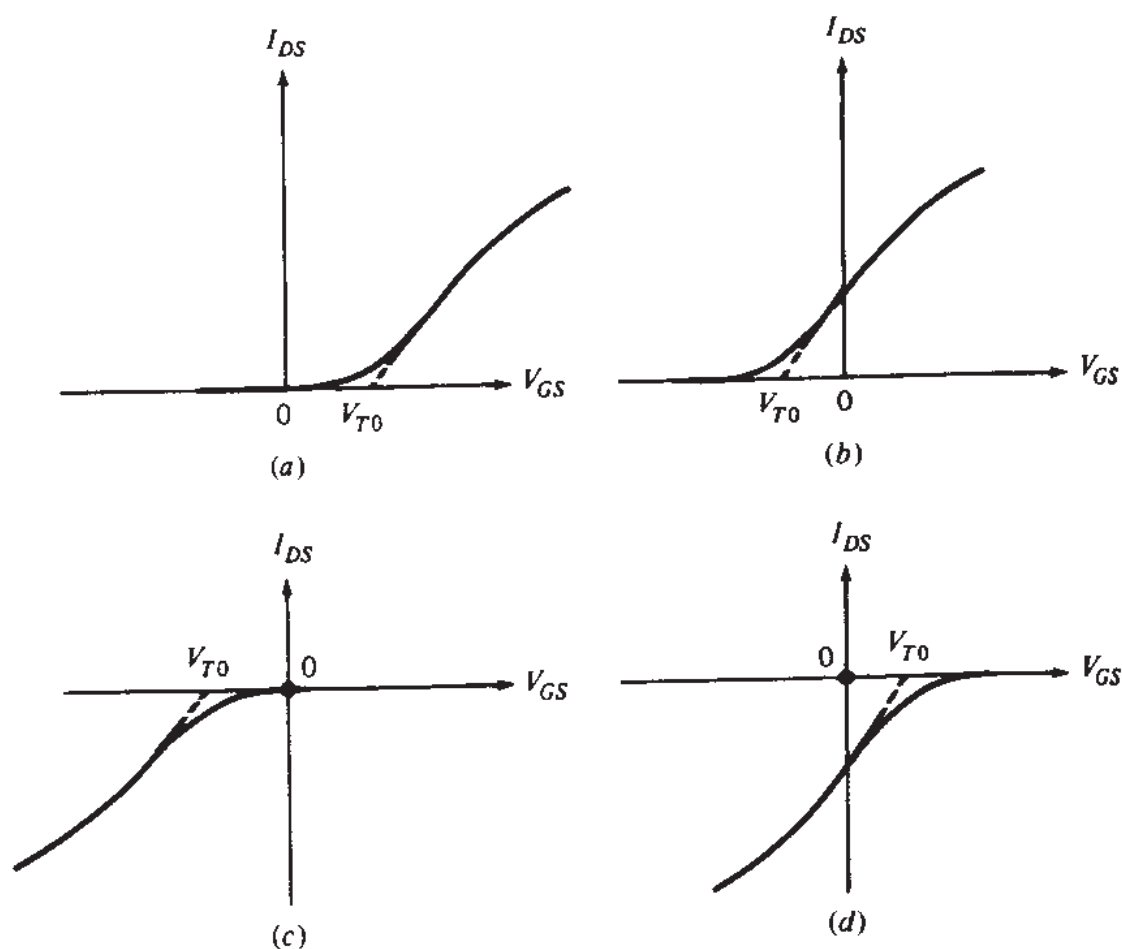


FIGURE 4.34

I_{DS} vs. V_{GS} for $V_{SB} = 0$ and very small $|V_{DS}|$. (a) n -channel enhancement device; (b) n -channel depletion device; (c) p -channel enhancement device; (d) p -channel depletion device.

The value of V_{T0} can, in principle, be set by a very shallow ion implantation during fabrication. In such a process, the device is bombarded with high-energy ions. If all such ions could end up at the oxide-semiconductor interface, their effect would be the same as that of Q'_o in (2.2.6). Hence, the value of V_{FB} could be adjusted, which, in turn, would adjust the value of V_{T0} in (4.5.33b). In practice, however, ion implantation cannot be that shallow. The spreading of ions into the substrate can create second-order effects, and then the resulting device cannot be modeled accurately as in this chapter. Ion-implanted devices are considered in Chap. 5.

4.15 MODEL PARAMETER VALUES, MODEL ACCURACY, AND MODEL COMPARISON

In the models we have developed, some parameters do not have an exact theoretical value. For example, such is the case with ϕ_0 , which represents the supposedly "pinned" value of the surface potential in strong inversion (with $V_{SB} = 0$). Since the surface potential is never exactly pinned but instead varies with bias, the best single value to be used for ϕ_0 will depend on the bias range we are interested in. The same is true for V_{T0} , which contains ϕ_0 in its definition. Even parameter values which have been specified "exactly" (e.g., the body effect coefficient γ) are actually exact only for the fictitious, idealized device corresponding to the simplifying assumptions we have made (e.g., that the substrate doping is "exactly" uniform). Thus, when a model is used to represent a real device, the parameter values that will result in minimum error between model predictions and experiment are not necessarily dictated by a simple theory. For this reason, what is usually done is to assign values to these parameters in such a way as to provide "best matching" between model predictions and measurements. In this process, the "theoretical" values for the parameters might simply play the role of an initial guess. Of course, what constitutes "best matching" is subject to interpretation. One might desire good matching for devices with $L = 0.5 \mu\text{m}$ and V_{DS} from 1 to 2 V; or, most likely, for devices of various lengths in a wide range of bias voltages. The sets of parameter values to be used in these two cases will not necessarily be the same, and the matching is likely to be better in the first case, assuming the same model is used. Also, if the same parameter appears in two different models (e.g., ϕ_0 appears in both the complete model and the simplified strong-inversion model of Sec. 4.5), different values for it might have to be used in each model for best results. For this reason direct comparison of model equations is difficult and should be taken with a grain of salt.

Let us offer an example of problems that can arise when model predictions are compared to experiment. Let us say that the weak-inversion model of (4.6.13) for given terminal voltages and process parameters predicts a current 4 times larger than the measured value. Is the model bad? Not necessarily. The problem may just be that the process parameters are not accurately known. This is notably true of V_{FB} , for example, which appears in the exponential term in (4.6.13) through V_M . With $n = 1.3$, a 50-mV error in the value of V_{FB} is enough to give 4 times the actual current! This, of course, happens because in weak inversion the $\log I_{DS}$ vs. V_{GS} curves (e.g., Fig. 4.26) are so steep that a slight shift horizontally corresponds to a large current

change. Note that, if I - V plots were compared rather than numbers, the above problem with V_{FB} could have been spotted immediately. One would see that the curves produced by the model had the same shape as the measured ones, only they were shifted somewhat horizontally. Thus, V_{FB} would be suspected and a new value would be tried for that parameter. In general, blind number comparison can be extremely misleading.

In the above example, the error in the value of V_{FB} could have been the result of inadequate knowledge of process parameters, for example oxide thickness or substrate doping. In other cases, though, V_{FB} or some other parameters may have to be given a wrong value on purpose, in order to make up for model deficiencies. This should be alarming. *If a model is good, the optimum parameter values which provide best fit to measurements should be close to the values expected from theory.* This point is discussed further in Chap. 10.

The determination of parameter values is done by using sophisticated automated systems devoted to data acquisition and parameter extraction. Under the central control of a computer, measurements such as I - V or C - V characteristics are performed on many devices of a certain type. The data obtained are then used as input to software which contains the algorithms for the determination of parameter values. Crucial in this process is the minimization of a certain error (e.g., mean square error) between the measured data and the values predicted by the model used. The selection of an appropriate error criterion depends on the application for which the model is intended. For example, for analog applications the error should involve small-signal parameters. Values for model parameters are extracted from I - V characteristic regions where such parameters have a strong effect. For example, if nonsaturation drain current values are plotted vs. V_{GS} for a very small V_{DS} , we obtain a curve of the form of Fig. 4.28; if, in the region of maximum slope of that curve, we draw a tangent and extrapolate to the V_{GS} axis, we obtain the value of V_T .† Parameter extraction is discussed in Chap. 10.

We shall take the opportunity here to emphasize that in this book V_T is the *extrapolated* threshold voltage, a term originating in Sec. 3.4.2 and also justified by the construction we have just described. Unfortunately, the term *threshold voltage* is used in the literature with at least three other meanings. Sometimes it is taken to denote the quantity $V_{FB} + 2\phi_F + \sqrt{2\phi_F + V_{SB}}$. This quantity is actually V_M , the value of V_{GS} at the lower limit of moderate inversion, as has been seen in Sec. 4.4. It is *different* from V_T in (4.5.32), simply because ϕ_0 is different from $2\phi_F$ (Sec. 2.5.2). Elsewhere, “threshold voltage” is used to imply what is sometimes called *constant current threshold voltage*. This is the value of V_{GS} needed to reach a set value of $I_{DS}/(W/L)$, which often happens to fall somewhere in the moderate-inversion region or even in the weak-inversion region. Finally, sometimes threshold voltage is taken to mean vaguely “the value of V_{GS} at which strong inversion begins.” The name “threshold

†More accurately, as can be verified from (4.5.37a), the V_{GS} axis intercept of that tangent gives $V_T + (a/2)V_{DS}$.

voltage" is often used indiscriminately for all these quantities, and sometimes during parameter extraction it is attempted to match calculated values of one of these "thresholds" (e.g., $V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F + V_{SB}}$) to measured values of another threshold (e.g., the extrapolated threshold). All this originated in the early days, when V_{GS} values of, say, 15 V were not uncommon; compared to these, an ambiguity of, say, 0.2 V in the value of threshold was not a problem. Today, however, when the trend is clearly toward low-voltage circuit operation, a distinction of the above quantities is important. In this book "threshold voltage" will always imply *extrapolated* threshold voltage unless stated otherwise. The extrapolation can be in terms of current, as in this section, or in terms of inversion layer charge, as in Fig. 3.2d.

REFERENCES

1. J. E. Lilienfeld, U.S. Patents 1,745,175 (1930) (filed Oct. 8, 1926), 1,877,140 (1932), 1,900,018 (1933).
2. O. Heil, British Patent 439,457 (filed and granted 1935).
3. W. Shockley and G. L. Pearson, "Modulation of conductance of thin films of semiconductors by surface charges," *Physical Review*, vol. 74, pp. 232–233, July 1948.
4. D. Kahng and M. M. Atalla, "Silicon-silicon dioxide field induced devices," *Solid-State Device Research Conference*, Pittsburgh, June 1960.
5. C.-T. Sah, "Evolution of the MOS transistor—From conception to VLSI," *Proceedings IEEE*, vol. 76, no. 10, pp. 1280–1326, October 1988.
6. H. K. J. Ihantola, "Design theory of surface field-effect transistor," Stanford Electronics Laboratories, *Technical Report No. 1161-I*, Stanford University, California, September 1961.
7. S. R. Hofstein and F. P. Heinman, "The silicon insulated-gate field effect transistor," *Proceedings of the IEEE*, vol. 51, pp. 1190–1202, September 1963.
8. H. K. J. Ihantola and J. L. Moll, "Design theory of a surface field-effect transistor," *Solid-State Electronics*, vol. 7, pp. 423–430, June 1964.
9. C. T. Sah, "Characteristics of the metal-oxide-semiconductor transistors," *IEEE Transactions on Electron Devices*, vol. ED-11, pp. 324–345, July 1964.
10. H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors," *Solid-State Electronics*, vol. 10, p. 927–937, 1966.
11. J. T. Wallmark and H. Johnson, *Field Effect Transistors: Physics, Technology and Applications*, Prentice-Hall, Englewood Cliffs, 1966.
12. C. T. Sah and H. C. Pao, "The effects of fixed bulk charge on the characteristics of metal-oxide semiconductor transistors," *IEEE Transactions on Electron Devices*, vol. ED-13, pp. 393–409, April 1966.
13. R. S. Cobbold, "MOS transistor as a four terminal device," *Electronics Letters*, vol. 2, pp. 189–190, June 1966.
14. J. A. Van Nielen and O. W. Memelink, "The influence of the substrate upon the D.C. characteristics of silicon MOS transistors," *Philips Research Reports*, vol. 22, pp. 55–71, February 1967.
15. M. B. Das, "Dependence of the characteristics of MOS transistors on the substrate resistivity," *Solid-State Electronics*, vol. 11, pp. 305–322, 1968.
16. A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley, New York, 1967.
17. M. B. Das, "Physical limitation of MOS structures," *Solid-State Electronics*, vol. 12, pp. 305–336, May 1969.
18. R. S. C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley-Interscience, New York, 1970.
19. P. Richman, *Characteristics and Operation of MOS Field Effect Devices*, McGraw-Hill, New York, 1967.

20. R. H. Crawford, *MOSFET in Circuit Design*, McGraw-Hill, New York, 1967.
21. H. Shichman and D. A. Hodges, "Modeling and simulation of insulated-gate field-effect transistor switching circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-3, pp. 285-289, September 1968.
22. D. Frohman-Bentckowsky and L. Vadasz, "Computer-aided design and characterization of digital MOS integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-4, pp. 57-64, April 1969.
23. J. E. Meyer, "MOS models and circuit simulations," *RCA Review*, vol. 32, pp. 42-63, March 1971.
24. G. Merckel, J. Borel, and N. Z. Cupcea, "An accurate large-signal MOS transistor model for use in computer-aided design," *IEEE Transactions on Electron Devices*, vol. ED-19, pp. 681-690, May 1972.
25. F. M. Klaassen, "A MOS model for computer-aided design," *Philips Research Reports*, vol. 31, pp. 71-83, 1976.
26. Y. Hayashi and Y. Tarui, "Exponential current in MOST-type devices and deterioration of reverse current in p-n junctions," Technical Group on Semiconductors and Semiconductor Devices, Institute of Electronics and Communications Engineers of Japan, *Technical Report SSD 67-6*, 1967.
27. A. A. Guzev, G. L. Kuryshv, and S. P. Sinita, "Investigation of carrier capture on the surface of silicon in a metal-insulator-semiconductor transistor," *Soviet Physics—Semiconductors*, vol. 4, pp. 1245-1249, February 1971.
28. R. M. Swanson and J. D. Meindl, "Ion-implanted complementary MOS transistors in low-voltage circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-7, pp. 146-153, April 1972.
29. M. B. Barron, "Low-level currents in insulated-gate field-effect transistors," *Solid-State Electronics*, vol. 15, pp. 293-302, 1972.
30. R. A. Stuart and W. Eccleston, "Leakage currents of MOS devices under depletion conditions," *Electronics Letters*, vol. 8, pp. 225-227, May 1972.
31. R. R. Troutman and S. N. Chakravarti, "Subthreshold characteristics of insulated-gate field-effect transistors," *IEEE Transactions on Circuit Theory*, vol. CT-20, pp. 659-665, November 1973.
32. R. J. Van Overstraeten, G. Declerck, and G. L. Broux, "Inadequacy of the classical theory of the MOS transistor operation in weak inversion," *IEEE Transactions on Electron Devices*, vol. ED-20, pp. 1150-1153, December 1973.
33. R. R. Troutman, "Subthreshold design consideration for insulated gate field-effect transistors," *IEEE Journal of Solid-State Circuits*, vol. SC-9, pp. 55-60, April 1974.
34. T. Masuhara, J. Etoh, and M. Nagata, "A precise MOSFET model for low-voltage circuits," *IEEE Transactions on Electron Devices*, vol. ED-21, pp. 363-371, June 1974.
35. R. J. Van Overstraeten, G. J. Declerck, and P. A. Muls, "Theory of MOS transistor in weak inversion—new method to determine the number of surface states," *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 282-288, May 1975.
36. R. R. Troutman, "Subthreshold slope for insulated gate field-effect transistors," *IEEE Transactions on Electron Devices*, pp. 1049-1051, November 1975.
37. E. Vittoz and J. Fellrath, "CMOS analog circuits based on weak inversion operation," *IEEE Journal of Solid-State Circuits*, vol. SC-12, pp. 224-231, June 1977.
38. G. W. Taylor, "Subthreshold conduction in MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-25, pp. 337-350, March 1978.
39. W. Fichtner and H. W. Potzl, "MOS modelling by analytical approximations. I. Subthreshold current and threshold voltage," *International Journal of Electronics*, vol. 46, pp. 33-55, 1979.
40. I. R. M. Mansour, "On the modeling of MOS devices," *Proceedings of the Third International Symposium on Network Theory*, Yugoslavia, pp. 705-713, 1975.
41. I. R. M. Mansour, "Improved modeling of MOS devices," *Proceedings of the European Conference on Circuit Theory and Design*, Italy, 1976.
42. Y. A. El-Mansy and A. R. Boothroyd, "A new approach to the theory and modeling of insulated-gate field-effect transistors," *IEEE Transactions on Electron Devices*, vol. ED-24, pp. 241-253, March 1977. See also related comments by J. R. Brews, *ibid.*, pp. 1369-1370, December 1977, and reply by Y. A. El-Mansy and A. R. Boothroyd, *ibid.*, vol. ED-25, pp. 393-394, March 1978.

43. G. Baccarani, M. Rudan, and G. Spadini, "Analytical i.g.f.e.t. model including drift and diffusion currents," *IEEE Journal on Solid-State and Electron Devices*, vol. 2, pp. 62–68, March 1978.
44. J. R. Brews, "A charge sheet model for the MOSFET," *Solid-State Electronics*, vol. 21, pp. 345–355, 1978.
45. F. Van de Wiele, "A long-channel MOSFET model," *Solid-State Electronics*, vol. 22, pp. 991–997, 1979.
46. J. R. Brews, "Physics of the MOS transistor," chap. 1 in *Silicon Integrated Circuits, Part A*, D. Kahng (editor), Applied Solid-State Science Series, Academic Press, New York, 1981.
47. R. F. Pierret and J. A. Shields, "Simplified long-channel MOSFET theory," *Solid-State Electronics*, vol. 26, pp. 143–147, 1983.
48. H. K. Lim and J. G. Fossum, "An analytic characterization of weak-inversion drift current in a long-channel MOSFET," *IEEE Transactions on Electron Devices*, vol. ED-30, pp. 713–715, June 1983.
49. C. Turchetti, "Relationships for the drift and diffusion components of the drain current in an MOS transistor," *Electronics Letters*, vol. 19, pp. 960–962, November 10, 1983.
50. M. Bagheri and C. Turchetti, "The need for an explicit model describing MOS transistors in moderate inversion," *Electronics Letters*, vol. 21, pp. 873–874, September 12, 1985.
51. P. P. Guebels and F. Van de Wiele, "A charge sheet model for small geometry MOSFET's," *Technical Digest*, International Electron Devices Meeting, pp. 211–214, Washington, 1981.
52. C. Turchetti and G. Masetti, "A CAD-oriented analytical MOSFET model for high-accuracy applications," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. CAD-3, pp. 117–122, 1984.
53. S. Yu, A. F. Franz, T. G. Mihran, "A physical parametric transistor model for CMOS circuit simulation," *IEEE Transactions on Computer-Aided Design*, vol. 7, pp. 1038–1052, October 1988.
54. H.-J. Park, P. K. Ko, and C. Hu, "A charge sheet capacitance model of short channel MOSFET's for SPICE," *IEEE Transactions on Computer-Aided Design*, vol. 10, pp. 376–389, March 1991.
55. A. R. Boothroyd, S. W. Tarasewicz, and C. Slaby, "MISNAN—A physically based continuous MOSFET model for CAD applications," *IEEE Transactions on Computer-Aided Design*, vol. 10, pp. 1512–1529, December 1991.
56. M. Miura-Mattausch, U. Feldmann, A. Rahm, M. Bollu, and D. Savignac, "Unified complete MOSFET model for analysis of digital and analog circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, pp. 1–7, January 1996.
57. N. Arora, *MOSFET Models for VLSI Circuit Simulation—Theory and Practice*, Springer-Verlag, Vienna, 1993.
58. S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.
59. C. C. McAndrew, private communication.
60. M. A. Maher and C. A. Mead, "A physical charge-controlled model for MOS transistors," pp. 211–229, in P. Lesleben (editor), *Advanced Research VLSI*, The MIT Press, Cambridge, Mass., 1987.
61. M. Bagheri and Y. Tsvitidis, "A small-signal dc-to-high-frequency nonquasistatic model for the four-terminal MOSFET valid in all regions of operation," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 2383–2391, November 1985.
62. B. Iñiguez and E. G. Moreno, "A physically based C_{∞} -continuous model for small-geometry MOSFET's," *IEEE Transactions on Electron Devices*, vol. 42, pp. 283–287, February 1995.
63. M. Bagheri, "Improving the non-quasistatic weak-to-strong-inversion four-terminal MOSFET model," *IEEE Transactions on Electron Devices*, vol. ED-34, no. 12, pp. 2558–2560, December 1987.
64. A. I. A. Cunha, M. C. Schneider, and C. G. Montoro, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electronics*, vol. 38, pp. 1945–1952, November 1995.
65. A. I. A. Cunha, *Um Modelo do Transistor MOS para Projecto de Cicuitos Integrados*, Ph.D. thesis, Universidade Federal de Santa Catarina, December 1996 (in Portuguese).
66. H. Oguey and S. Cserveny, "Modèle du transistor MOS valable dans un grand domaine de courants," *Sonderdruck aus dem Bulletin des SEV/VSE*, vol. 73, pp. 113–116, 1982.

67. E. A. Vittoz, "Micropower Techniques," in *Design of Analog-Digital VLSI Circuits for Telecommunications and Signal Processing*, J. E. Franca and Y. Tsividis (editors), Prentice-Hall, Englewood Cliffs, N.J., 1994.
68. C. C. Enz, "High precision CMOS micropower amplifiers," Ph.D. thesis no. 802, École Polytechnique Fédérale de Lausanne, Switzerland, 1989.
69. C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog Integrated Circuits and Signal Processing*, vol. 8, pp. 83–114, July 1995.
70. C.-K. Park, C.-Y. Lee, K. Lee, B.-J. Moon, Y. H. Byun, and M. Shur, "A unified current-voltage model for long-channel nMOSFET's," *IEEE Transactions on Electron Devices*, vol. 38, pp. 399–406, February 1991.
71. K. Lee, M. Shur, T. A. Fjeldly, and T. Ytterdal, *Semiconductor Device Modeling for VLSI*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
72. A. I. A. Cunha, O. C. Gouveia-Filho, M. C. Schneider, and C. Galup-Montoro, "A current-based model for the MOS transistor," *Proceedings 1997 International Symposium on Circuits and Systems*, pp. 1608–1611, Hong Kong, June 1997.
73. A. Vladimirescu and S. Liu, "The simulation of ICs using SPICE-2," University of California—Berkeley, Electronics Research Laboratory, Memorandum M80/7, 1980.
74. P. G. A. Jespers, C. Jussere, and Y. Leduc, "A fast sample and hold charge-sensing circuit for photodiode arrays," *IEEE Journal of Solid-State Circuits*, vol. SC-12, pp. 232–237, June 1977.
75. H. Wallinga and K. Bult, "Design and analysis of CMOS analog signal processing circuits by means of a graphical MOST model," *IEEE Journal of Solid-State Circuits*, vol. 24, p. 672, June 1989.
76. Y. Tsividis and G. Masetti, "Problems in precision modeling of the MOS transistor for analog applications," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. CAD-3, pp. 72–79, January 1984.
77. K. A. Valiev, A. N. Karmazinskii, and A. M. Korolev, *Digital Integrated Circuits Using MOS Transistor*, Soviet Radio, Moscow, 1971.
78. J. A. Geurst, "Calculation of high frequency characteristics of thin film transistors," *Solid-State Electronics*, vol. 8, pp. 88–90, 1965.
79. J. J. Paulos and D. A. Antoniadis, "Limitations of quasi-static capacitance models for the MOS transistor," *IEEE Electron Device Letters*, vol. EDL-4, pp. 221–224, July 1983.
80. C. C. McAndrew, B. K. Bhattacharyya, and O. Wing, "A single-piece C_{ox} -continuous MOSFET model including subthreshold conductance," *IEEE Electron Device Letters*, vol. 12, pp. 565–567, October 1991.
81. J. A. Power, and W. A. Lane, "An enhanced SPICE MOSFET model suitable for analog applications," *IEEE Transactions on Computer-Aided Design*, vol. 11, pp. 1418–1425, November 1992.
82. N. D. Arora, R. Rios, C.-L. Huang, and K. Raol, "PCIM: a physically based continuous short-channel IGFET model for circuit simulation," *IEEE Transactions on Electron Devices*, vol. 41, pp. 988–997, June 1994.
83. F. M. Klaassen, "Review of physical models for MOS transistors," in *Process and Device Modeling for Integrated Circuit Design*, F. Van de Wiele, W. L. Engl, and P. G. Jespers (editors), Noordhoff, Leyden, The Netherlands, 1977.
84. H. C. Poon, private communication, 1977.
85. F. M. Klaassen, "MOS device modeling," in *Design of VLSI Circuits for Telecommunications*, Y. Tsividis and P. Antognetti (editors), Prentice-Hall, Englewood Cliffs, N.J., 1985.
86. G. Merckel, "CAD models of MOSFETS," in *Process and Device Modelling for Integrated Circuit Design*, F. Van de Wiele, W. L. Engl, and P. G. Jespers (editors), Noordhoff, Leyden, The Netherlands, 1977.
87. H. I. Hanafi, L. H. Camnitz, and A. J. Dally, "An accurate and simple MOSFET model for computer-aided design," *IEEE Journal of Solid-State Circuits*, vol. SC-17, pp. 882–891, October 1982.

88. S. Liu and L. W. Nagel, "Small-signal MOSFET models for analog circuit design," *IEEE Journal of Solid-State Circuits*, vol. SC-17, pp. 983–998, December 1982.
89. P. Antognetti, D. D. Caviglia, and E. Profumo, "CAD model for threshold and subthreshold conduction in MOSFETs," *IEEE Journal of Solid-State Circuits*, vol. SC-17, pp. 454–458, June 1982.
90. G. T. Wright, "Simple and continuous MOSFET models for the computer-aided design of VLSI," *IEE Proceedings*, vol. 132, part I, pp. 187–194, August 1985.
91. Y. Tsividis, K. Suyama, and K. Vavelidis, "A simple 'Reconciliation' MOSFET model valid in all regions," *Electronics Letters*, vol. 31, March 1995, pp. 506–508.
92. P. Rossel, "Influence de la réduction de mobilité due au champ transversal sur les caractéristiques des m.o.s.t.," *Electronics Letters*, vol. 5, pp. 604–605, 1969.
93. S. A. Schwarz and S. E. Russek, "Semi-empirical equations for electron velocity in silicon—Part II: MOS inversion layer," *IEEE Transactions on Electron Devices*, vol. ED-30, pp. 1634–1639, December 1983.
94. A. G. Sabnis and J. T. Clemens, "Characterization of the electron mobility in the inverted <100> Si surface," *Technical Digest, IEEE International Electron Devices Meeting*, pp. 18–21, Washington, 1979.
95. J. T. C. Chen and R. S. Muller, "Carrier mobilities at weakly inverted silicon surfaces," *Journal of Applied Physics*, vol. 45, p. 828, 1974.
96. J. R. Brews, "Carrier-density fluctuations and the IGFET mobility near threshold," *Journal of Applied Physics*, vol. 46, p. 2193, 1975.
97. C. G. Sodini, T. Ekstedt, and J. L. Moll, "Charge accumulation and mobility in thin dielectric MOS transistors," *Solid-State Electronics*, vol. 25, pp. 833–841, September 1982.
98. S. C. Sun and J. D. Plummer, "Electron mobility in inversion and accumulation layers on thermally oxidized silicon surface," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 1497–1508, 1980.
99. T. Ando, A. B. Fowler, and F. Stern, "Electronic properties of two-dimensional systems," *Reviews of Modern Physics*, vol. 54, pp. 437–672, 1982.
100. K. Y. Fu, "Mobility degradation due to the gate field in the inversion layer of MOSFETs," *IEEE Electron Device Letters*, vol. EDL-3, pp. 292–293, 1982.
101. M.-S. Liang, J. Y. Choi, P.-K. Ko, and C. Hu, "Inversion-layer capacitance and mobility of very thin gate-oxide MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-33, p. 409, March 1986.
102. B. Majkusiak and A. Jakubowski, "The dependence of MOSFET surface carrier mobility on gate-oxide thickness," *IEEE Transactions on Electron Devices*, vol. ED-33, p. 1717, November 1986.
103. N. D. Arora and G. Sh. Gildenblat, "A semi-empirical model of the MOSFET inversion layer mobility for low-temperature operation," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 89–93, 1987.
104. J. W. Watt and J. D. Plummer, "Universal mobility-field curves for electrons and holes in MOS inversion layers," *Proceedings Symposium on VLSI Technology*, pp. 81–82, 1987.
105. A. J. Walker and P. H. Woerlee, "Mobility model for silicon inversion layers," *EESDERC 1987, Technical Digest*, pp. 667–670.
106. A. Hiroki, S. Odanaka, K. Ohe, and H. Esaki, "A mobility model for submicrometer MOSFET device simulations," *IEEE Electron Device Letters*, vol. EDL-8, pp. 231–233, May 1987.
107. W. M. Soppa and H.-G. Wagemann, "Investigation and modeling of the surface mobility of MOSFETs from -25 to $+150^{\circ}\text{C}$," *IEEE Transactions on Electron Devices*, vol. 35, p. 970, July 1988.
108. J. A. Wikstrom and C. R. Viswanathan, "Lateral nonuniformities and the MOSFET mobility step near threshold," *IEEE Transactions on Electron Devices*, vol. 35, pp. 2378–2383, December 1988.
109. S.-W. Lee, "Universality of mobility-gate field characteristics of electrons in the inversion charge layer and its application in MOSFET modeling," *IEEE Transactions on Computer-Aided Design*, vol. 8, pp. 724–730, July 1989.
110. D. T. Amm, H. Mingam, P. Delpech, and T. T. D'ouville, "Surface mobility in n^+ and p^+ doped polysilicon gate PMOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-36, pp. 963–967, 1989.

111. H. Shin, A. F. Tasch, Jr., C. M. Mazier, and S. K. Banerjee, "A new approach to verify and derive a transverse-field-dependent mobility model for electrons in MOS inversion layers," *IEEE Transactions on Electron Devices*, vol. 36, p. 1117, June 1989.
112. D. S. Jeon and D. E. Burk, "MOSFET electron inversion layer mobilities—a physically based semi-empirical model for a wide temperature range," *IEEE Transactions on Electron Devices*, vol. 36, p. 1456, August 1989.
113. J. S. Kang, D. K. Schroder, and A. R. Alvarez, "Effective and field-effect mobilities in Si MOSFETS," *Solid-State Electronics*, vol. 32, pp. 679–681, 1989.
114. G. M. Yeric, A. F. Tasch, and S. K. Banerjee, "A universal MOSFET mobility degradation model for circuit simulation," *IEEE Transactions on Computer-Aided Design*, vol. 9, pp. 1123–1126, October 1990.
115. C.-L. Huang and G. S. Gildenblat, "Measurements and modeling of the n-channel MOSFET inversion layer mobility and device characteristics in the temperature range 60–300 K," *IEEE Transactions on Electron Devices*, vol. 37, pp. 1289–1300, May 1990.
116. D. B. M. Klaassen, "A unified mobility model for device simulation," *Digest, International Electron Devices Meeting*, pp. 357–360, December 1990.
117. P. Habas, "A physics based analytical MOSFET model with accurate field dependent mobility," *Solid-State Electronics*, vol. 33, pp. 923–933, 1990.
118. H. C. de Graaff and F. M. Klaassen, *Compact Transistor Modelling for Circuit Design*, Springer-Verlag, Vienna, 1990.
119. V. M. Agostinelli, Jr., H. Shin, and A. F. Tasch, Jr., "A comprehensive model for inversion layer hole mobility for simulation of submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, vol. 38, p. 151, January 1991.
120. H. Shin, G. M. Yeric, A. F. Tasch, and C. M. Maziar, "Physically-based models for effective mobility of electrons in MOS inversion layers," *Solid-State Electronics*, vol. 34, pp. 545–552, 1991.
121. M. Shirahata, H. Kusano, N. Kotani, S. Kusanoki, and Y. Akasaka, "A mobility model including the screening effect in MOS inversion layer," *IEEE Transactions on Computer-Aided Design*, vol. 11, pp. 1114–1119, September 1992.
122. V. M. Agostinelli, Jr., G. M. Yeric, and A. F. Tasch, Jr., "Universal MOSFET hole mobility degradation models for circuit simulation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 12, pp. 439–445, March 1993.
123. K. Lee, M. Shur, T. A. Fjeldly, and T. Ytterdal, *Semiconductor Device Modeling for VLSI*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
124. C.-L. Huang, J. V. Faricelli, and N. D. Arora, "A new technique for measuring MOSFET inversion layer mobility," *IEEE Transactions on Electron Devices*, vol. 40, pp. 1134–1139, June 1993.
125. H.-S. Wong, "Universal effective mobility of empirical local mobility models for n- and p-channel silicon MOSFETs," *Solid-State Electronics*, vol. 36, pp. 179–188, 1993.
126. C. Yue, M. Agostinelli, Jr., G. M. Yeric, and A. F. Tasch, "Improved universal MOSFET electron mobility degradation models for circuit simulation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 12, pp. 1542–1546, October 1993.
127. S.-i. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFET's: Part I—Effects of substrate impurity concentration," *IEEE Transactions on Electron Devices*, vol. 41, pp. 2357–2362, December 1994.
128. S.-i. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFET's: Part II—Effects of surface orientation," *IEEE Transactions on Electron Devices*, vol. 41, pp. 2363–2368, December 1994.
129. K. Yang, R. C. Meitzler, and A. G. Andreou, "A model for MOS effective channel mobility with emphasis in the subthreshold and transition region," *Proceedings IEEE 1994 International Symposium on Circuits and Systems*, vol. 1, pp. 431–434.
130. A. Emrani, G. Ghibaudo, and F. Balestra, "On the universal electric field dependence of the electron and hole effective mobility in MOS inversion layers," *Solid-State Electronics*, vol. 37, pp. 111–113, January 1994.

131. F. Gámiz, J. Banqueri, J. E. Carceller, and J. A. L.-Villanueva, "Effects on bulk-impurity and interface-charge on the electron mobility in MOSFETs," *Solid-State Electronics*, vol. 38, pp. 611–614, March 1995.
132. F. Gámiz, J. A. L.-Villanueva, J. Banqueri, J. E. Carceller, and P. Cartujo, "Universality of electron mobility curves in MOSFETs: a Monte Carlo study," *IEEE Transactions on Electron Devices*, vol. 42, pp. 258–265, February 1995.
133. F. Gámiz, J. A. L.-Villanueva, J. Banqueri, and J. E. Carceller, "Influence of the oxide-charge distribution profile on electron mobility in MOSFET's," *IEEE Transactions on Electron Devices*, vol. 42, pp. 999–1004, May 1995.
134. J. Banqueri, J. A. L.-Villanueva, F. Gámiz, J. E. Carceller, E. L.-Tamayo, and M. Lozano, "A procedure for the determination of the effective mobility in an *N*-MOSFET in the moderate inversion region," *Solid-State Electronics*, vol. 39, pp. 875–883, June 1996.
135. J. R. Hauser, "Extraction of experimental mobility data for MOS devices," *IEEE Transactions on Electron Devices*, vol. 43, pp. 1981–1988, November 1996.
136. S. Villa, A. L. Lacaita, L. Perron, R. Bez, "Effective mobility in heavily doped *n*-MOSFET's: measurements and models," *Digest, International Electron Devices Meeting*, pp. 395–398, 1996.
137. S. Jallepalli, W.-K. Shih, J. D. Bude, M. R. Pinto, C. M. Mazier, and A. F. Tasch, Jr., "Understanding the differences in the effective-field dependence of electron and hole inversion layer mobilities," *Digest, International Electron Devices Meeting*, pp. 391–394, December 1996.
138. B. Cheng and J. Woo, "A temperature-dependent MOSFET inversion layer carrier mobility model for device and circuit simulation," *IEEE Transactions on Electron Devices*, vol. 44, pp. 343–345, February 1997.
139. M. H. White, F. Van de Wiele, and J. P. Lambot, "High-accuracy models for computer-aided design," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 899–906, May 1980.
140. G. Giralt, B. Andre, J. Simonne, and D. Esteve, "Influence de la température sur les dispositifs semiconducteurs du type M.O.S.," *Electronics Letters*, vol. 1, pp. 185–186, September 1965.
141. R. Wong, J. Dunkley, T. A. DeMassa, and J. F. Jelsma, "Threshold voltage variations with temperature in MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-18, p. 386, 1971.
142. S. K. Tewksbury, "N-channel enhancement-mode MOSFET characteristics from 10 to 300 K," *IEEE Transactions on Electron Devices*, vol. ED-28, pp. 1519–1529, December 1981.
143. F. Shoucair, W. Hwang, and P. Jain, "Electrical characteristics of large scale integration (LSI) MOSFETs at very high temperatures," *Microelectronics and Reliability*, vol. 24, part I, pp. 465–485, part II, pp. 487–510, 1984.
144. F. M. Klaassen and W. Hes, "On the temperature coefficient of the MOSFET threshold voltage," *Solid-State Electronics*, vol. 29, pp. 787–789, 1986.
145. G. S. Gildenblat and C.-L. Huang, "N-channel MOSFET model for the 60–300-K temperature range," *IEEE Transactions on Computer-Aided Design*, vol. 10, pp. 512–518, April 1991.
146. F. Balestra and G. Ghibaud, "Brief review of the MOS device physics for low temperature electronics," *Solid-State Electronics*, vol. 37, pp. 1967–1975, December 1994.

PROBLEMS

- 4.1. Plot the total drain current, the component due to drift, and the component due to diffusion, using relations from Sec. 4.3.1, for an *n*-channel transistor with $N_A = 5 \times 10^{17} \text{ cm}^{-3}$, $t_{\text{ox}} = 60 \text{ \AA}$, $V_{FB} = -1 \text{ V}$, $\mu = 600 \text{ cm}^2/(\text{V} \cdot \text{s})$, $W = L = 2 \text{ }\mu\text{m}$, biased with $V_{SB} = 1 \text{ V}$ and $V_{DB} = 3 \text{ V}$, for V_{GB} between 0 and 3 V. Use a log current axis. Identify the three regions of inversion on the plot, and evaluate the percentages of the current due to drift and due to diffusion at the limit points between weak and moderate inversion and between moderate and strong inversion.

- 4.2. As explained in Sec. 4.3.1, in using (4.3.17) in weak inversion a small error in each of ψ_{sL} and ψ_{s0} can cause a large relative error in I_{DS2} , since ψ_{sL} and ψ_{s0} have nearly equal values. Show that if $\psi_{sL} - \psi_{s0}$ in (4.3.17) is substituted by the difference of the right-hand sides of (4.3.18a) and (4.3.18b), an expression much more immune to error results.⁵² Explain why the new expression obtained may not be acceptable for use with (4.3.8) in strong inversion if ψ_{s0} and ψ_{sL} are not known accurately. (See also Prob. 4.4.)
- 4.3. Compare the simplified charge sheet model of (4.3.39) and (4.3.40) to the complete charge sheet model of Sec. 4.3.1, by constructing current plots for the parameters and voltage values specified in Prob. 4.1.
- 4.4. Consider the device of Prob. 4.1 with $V_{SB} = 1$ V, $V_{DB} = 3$ V, and V_{GB} corresponding to 0.2 V below the upper limit of weak inversion. Find ψ_{s0} and ψ_{sL} from (4.3.18) to very good accuracy (e.g., to eight significant digits) by using a computer or calculator with an equation-solving ability. Then, find I_{DS2} from (4.3.17) and from the expression of Prob. 4.2, assuming ψ_{s0} and ψ_{sL} are known to only five, four, three, and two significant digits each. Which expression is more immune to limited accuracy? Explain the reasons by using specific numbers from the computation.
- 4.5. Plot the inversion layer and depletion region charges per unit area vs. distance from the source for the device of Prob. 4.1, with $V_{SB} = 1$ V, $V_{DB} = 2$ V, and $V_{GB} = 3$ V.
- 4.6. For the device of Prob. 4.1, indicate the regions of inversion
 (a) In the manner of Fig. 4.12, for $V_{SB} = 0$ V and 1 V.
 (b) In the manner of Fig. 4.14, for $V_{GB} = 2$ V and 3 V.
- 4.7. This problem deals with various ways of defining regions of operation.
 (a) Without restricting V_{DS} to positive values, define $V = \min(V_{SB}, V_{DB})$ and give general definitions for the regions of inversion as described in Table 4.1, in terms of V and V_{GB} .
 (b) Repeat, in terms of V and V_{GA} , where $V_{GA} = \max(V_{GS}, V_{GD})$.
- 4.8. Starting from (4.5.2a), prove (4.5.2b), as well as the following equation:

$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \left\{ \frac{1}{2} \left[(V_{GS} - V_{FB} - \phi_0)^2 - (V_{GD} - V_{FB} - \phi_0)^2 \right] - \frac{2}{3} \gamma \left[(V_{DB} + \phi_0)^{3/2} - (V_{SB} + \phi_0)^{3/2} \right] \right\}$$

- 4.9. Starting from (4.5.2), show that the nonsaturation strong-inversion current can be written in terms of V_{GS} , V_{DS} , and V_{SB} as follows:

$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \left\{ (V_{GS} - V_{FB} - \phi_0) V_{DS} - \frac{1}{2} V_{DS}^2 - \frac{2}{3} \gamma \left[(\phi_0 + V_{SB} + V_{DS})^{3/2} - (\phi_0 + V_{SB})^{3/2} \right] \right\}$$

and find an expression for V'_{DS} , the value of V_{DS} at the boundary between nonsaturation and saturation.

- 4.10. For the device of Prob. 4.1, plot I_{DS} vs. V_{DB} (0 to 3 V) in strong inversion, for V_{GB} values of 2 and 3 V, assuming $V_{SB} = 0$, and for the following models:
 (a) The complete model of (4.5.2)

(b) The simplified model of (4.5.16)

(c) The simplified model of (4.5.23)

along with their saturation extensions. Comment on the agreement of (b) and (c) with (a).

- 4.11. For the device of Prob. 4.1 with $V_{SB} = 0$ V and $V_{GB} = 3$ V (strong inversion), plot I_{DS} vs. V_{DS} , from $V_{DS} = 0$ to $V_{DS} = 3$ V, using (a) the general model of Sec. 4.3.1 and (b) the strong inversion model of (4.5.13) and (4.5.2). Discuss the effect of the value of ϕ_0 in (b).
- 4.12. Consider the bracketed quantity containing the $3/2$ powers in Prob. 4.9. Starting from an expansion of this quantity into a Taylor series around $V_{DS} = 0$, derive (4.5.31). Prove that the value to be used for α should be less than α_1 in (4.5.41).
- 4.13. Show that, for (4.5.28) to correspond to line a in Fig. 4.16, α_1 should have the value given in (4.5.41).
- 4.14. For the transistor in Prob. 4.1, assume $V_{SB} = 0$ V and plot I_{DS} vs. V_{DS} for $V_{GS} = 2$ and 3 V, using (a) (4.5.13) with (4.5.2) and (b) (4.5.37) for each of the five values of α considered in Sec. 4.5.3. Comment on the accuracy obtained for each value of α . Can this accuracy be improved if α is allowed to be chosen at will? What value would you choose?
- 4.15. Show that the assumptions made in order to derive the simplified strong-inversion model of Sec. 4.5.3 lead to the conclusion that $Q'_I = 0$ when $V_{DS} = V'_{DS}$. Comment on the validity of this conclusion.
- 4.16. Show that the simplified source-referenced strong-inversion model of Sec. 4.5.3 is related to the symmetric models of Sec. 4.5.2. Start by writing (4.5.23) in terms of V_{GS} , and show that the resulting equation can be placed in the form of (4.5.37a), with V_T and α defined accordingly.
- 4.17. Prove (4.5.49).
- 4.18. For the simplified strong-inversion model of Sec. 4.5.3 and for the device of Prob. 4.1 with $\alpha = \alpha_3$, $V_{SB} = 1$ V, $V_{GS} = 3$ V, and $V_{DS} = V'_{DS} - 0.6$ V, plot (a) Q'_B and Q'_I as functions of V_{CB} , and (b) V_{CB} , Q'_B , and Q'_I as functions of the distance from the source.
- 4.19. For the device of Prob. 4.1 in weak inversion, plot (a) I_{DS} vs. V_{DS} , with $V_{SB} = 0$, for V_{DS} between 0 and 1 V, and for V_{GS} values of $V_M - 50$ mV and $V_M - 100$ mV. Use (4.6.6)–(4.6.7); (b) $\log I_{DS}$ vs. V_{GS} with $V_{DS} = 1$ V, for $V_{SB} = 0$ V and $V_{SB} = 2$ V.
- 4.20. Plot $\log I_{DS}$ vs. V_{GS} , for values as stated in Prob. 4.19, using (4.6.6)–(4.6.7) and (4.6.13). Compare the results. Can the agreement be improved if the values of V_M and n are slightly modified?
- 4.21. Show that a simple approximate expression for the current in weak inversion is

$$I_{DS} = \frac{W}{L} I'_X e^{(V_{GS} - V_X)/(n\phi_t)} (1 - e^{-V_{DS}/\phi_t})$$

where V_X is the value of V_{GS} corresponding to a surface potential of $1.5\phi_F + V_{SB}$, and

$$I'_X = \mu C'_{ox} \phi_t^2 \frac{\gamma}{2\sqrt{1.5\phi_F + V_{SB}}} e^{-0.5\phi_F/\phi_t}$$

Propose an appropriate expression for n . Comment on the accuracy of this equation in comparison to (4.6.13), and show that both equations can provide similar results if the values of the parameters in one of them are slightly modified.

- 4.22.** Show that (4.3.17) in weak inversion can be reduced to (4.6.6)–(4.6.7) as follows: Replace $\psi_{sL} - \psi_{s0}$ in (4.3.17) by the difference of the right-hand sides of (4.3.18b) and (4.3.18a). In this difference expand the square roots of the form $\sqrt{\psi_s + \xi}$ in a Taylor series around $\xi = 0$ and keep only the first two terms (justify this); in the resulting expression for I_{DS2} , use $\psi_{s0} \approx \psi_{sL} \approx \psi_{sa}$.
- 4.23.** Plot I_{DS} in nonsaturation vs. V_{GS} with $V_{SB} = 0$, $V_{DS} = 10$ mV, and V_{GS} between 1.5 and 3 V, assuming $(W/L)\mu C'_{ox} = 14 \mu\text{A/V}^2$, $N_A = 5 \times 10^{17} \text{ cm}^{-3}$, $t_{ox} = 60 \text{ \AA}$, and $V_{FB} = -1$ V. Now repeat, assuming a mobility dependence on V_{GS} as given by (4.10.20), with $\theta = 0.2 \text{ V}^{-1}$. Can V_T be obtained from the second plot by extrapolating to $I_{DS} = 0$ from the point of maximum slope?
- 4.24.** For the device of Prob. 4.23, plot I_{DS} vs. V_{DS} with V_{DS} between 0 and 3 V, for $V_{GS} = 2$ and 3 V.
- (a) Assume a constant mobility.
 - (b) Assume a V_{GS} -dependent mobility as given by (4.10.20) with $\theta = 0.2 \text{ V}^{-1}$.
- 4.25.** Rewrite the framed equations in this chapter for p -channel transistors.

CHAPTER 5

MOS TRANSISTORS WITH ION- IMPLANTED CHANNELS

5.1 INTRODUCTION

In the previous chapter, we assumed that the transistors we considered had a uniform substrate, as indicated in Fig. 5.1*a*. In real devices this assumption does not hold; the substrate doping is locally changed in the region between the source and the drain, as shown in Fig. 5.1*b*. This is achieved through ion implantation, a process in which the substrate is bombarded with ions during fabrication. The effective substrate doping concentration is changed in the areas where these ions land, and thus the substrate concentration varies as one goes from the surface toward the bulk.[†] A main aim of this is to set the threshold voltage of transistors to a desired value, but channel implantation is also used to modify other characteristics (notably the punchthrough behavior).

Since its early application to MOS transistors,¹⁻⁴ ion implantation has been used extensively and is currently a standard part of device fabrication.⁵⁻⁸ Ion implantation is used for more than doping the channel; for example, in Chap. 1 we encountered its use in forming the source and drain regions.² nMOS devices like the ones in Fig. 5.1 are formed by using ion implantation to dope the source and drain regions n^+ ; during this

[†]Even for unimplanted channels the assumption of a uniform substrate is not strictly correct, because of some impurity redistribution which takes place during oxide growth.

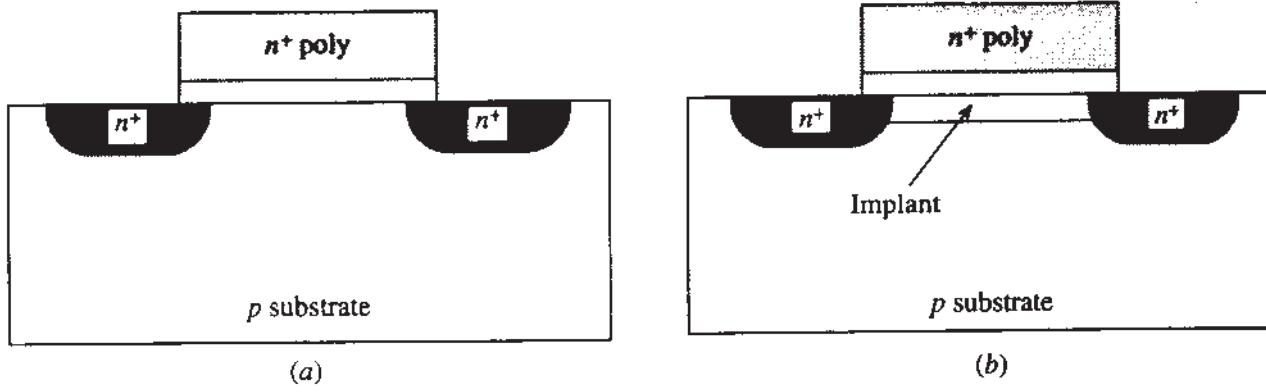


FIGURE 5.1
nMOS transistors. (a) With unimplanted channel; (b) with implanted channel.

process, the polysilicon gate acts as a mask, preventing the ions from landing in the channel area. In this process, the gate itself receives the ions, and is thus doped n^+ as indicated in the figure. However, in this chapter we will concentrate on the effects of ion implantation of the *channel* region.

The characteristics of devices with ion-implanted channels have been the subject of extensive studies.⁹⁻⁹⁶ The purpose of this chapter is to present some representative models which are simple enough to provide manageable analytical results, and to predict most of the important effects associated with implanted channels. Our quantitative results will be limited to drift currents, for which most of the analytical results in the literature have been obtained.

While a general analysis of devices with ion-implanted channels is possible, we prefer to introduce the principles involved by considering the typical use of this process in modern device fabrication. Thus, ion implantation is mainly used for making (1) enhancement nMOS devices, (2) depletion nMOS devices, and (3) enhancement pMOS devices. We will consider each case separately.

5.2 ENHANCEMENT nMOS TRANSISTORS

5.2.1 Preliminaries

Let us first consider the unimplanted nMOS device of Fig. 5.1a. The 0 - V_{SB} threshold voltage, V_{T0} , is given by (2.5.27), repeated here:

$$V_{T0} = \phi_{MS} - \frac{Q'_o}{C'_{ox}} + \phi_0 - \frac{Q'_B}{C'_{ox}} \quad (5.2.1)$$

where Q'_B is the depletion region charge in strong inversion, given by (2.5.25):

$$Q'_B = -\sqrt{2q\epsilon_s N_A \phi_0} \quad (5.2.2)$$

This equation can thus be written as in (2.5.28):

$$V_{T0} = V_{FB} + \phi_0 + \gamma\sqrt{\phi_0} \quad (5.2.3)$$

with

$$V_{FB} = \phi_{MS} - \frac{Q'_0}{C'_{ox}} \quad (5.2.4)$$

and

$$\gamma = \frac{\sqrt{2q\epsilon_s N_A}}{C'_{ox}} \quad (5.2.5)$$

If V_{SB} is not 0, the value of the threshold voltage V_T is changed from the above value due to the body effect; it is now given by (4.5.32) or (4.5.33), repeated here:

$$V_T = V_{FB} + \phi_0 + \gamma\sqrt{\phi_0 + V_{SB}} \quad (5.2.6a)$$

$$= V_{T0} + \gamma(\sqrt{\phi_0 + V_{SB}} - \sqrt{\phi_0}) \quad (5.2.6b)$$

For the enhancement-mode devices used in digital circuits, the value of V_{T0} should be large enough to make sure the device current will be negligible with $V_{GS} = 0$ (for which the device will be in weak inversion). This is needed in order to keep the static power dissipation of large digital circuits low, or to make sure that the charges used for capacitive storage in dynamic memories will not leak out. At the same time, V_{T0} should not be too high, so that with the maximum possible value of V_{GS} , large currents can be provided to charge parasitic capacitances fast (as required for high-speed operation). The maximum possible value of V_{GS} is normally equal to the power supply voltage, and the latter is only a few volts in modern technologies (Chap. 6). Thus, a compromise value must be chosen for V_{T0} , say 0.4 V.

To see what values of V_{T0} can be expected from the above equations, one can estimate V_{FB} in (5.2.4) as shown in Sec. 2.2. For modern devices Q'_0 is small and C'_{ox} is large (since the oxide is as thin as 30 or 40 Å); thus, Q'_0/C'_{ox} can be neglected to first order. The value of ϕ_{MS} depends to some degree on the substrate doping (Secs. 2.2 and 1.4), but not strongly; for n^+ poly gates, a typical value for ϕ_{MS} is -0.85 V. If we use for ϕ_0 a typical value of 0.75 V, we can see that $V_{FB} + \phi_0$ in (5.2.3) is about -0.1 V. Thus for this equation to give a sufficiently positive V_{T0} , a large value of γ will be needed. In modern technologies, where oxides are thin and thus C'_{ox} is large, this means that the substrate doping will have to be high, as follows from (5.2.5).

The use of a large γ , though, creates problems. As follows from Sec. 4.6, such a value results in a large value for the quantity n , and this causes a small value for the slope of $\log I_{DS}$ vs. V_{GS} in weak inversion. Thus, when V_{GS} is made 0, it may not be possible to turn the device adequately off in digital circuits. In analog applications, a large γ causes large variations of V_T when V_{SB} changes [see (5.2.6)], which can cause complications in circuit design. Finally, the high substrate doping needed in this solution causes large junction capacitances (Sec. 1.5), which reduces circuit speed (Chaps. 7 to 9).

Thus using a large N_A is not an attractive way to set the value of V_{T0} . What can be sought instead is to implant acceptor ions (e.g., boron) very close to the surface. Ionized acceptor atoms have a negative charge, which would be effectively added to Q'_o in (5.2.4), thus changing the effective V_{FB} and providing the required threshold increase. On the other hand, the substrate concentration would not be increased at points below the p -type implant where most of the depletion region is located. Thus the body effect coefficient and the junction capacitances would be kept small.

Unfortunately the location of the implant close to the surface, although optimal for threshold adjustment, is not optimal for handling a different problem encountered in short-channel devices. As will be discussed in Chap. 6, in such devices the depletion region lengths become significant in comparison to the channel length, resulting in several two-dimensional effects including a form of breakdown known as *punchthrough*. To prevent punchthrough, the depletion region widths should be contained at points between source and drain below the surface at a depth slightly smaller than the source and drain junction depth, where punchthrough currents can flow. This can be achieved by using ion implantation to selectively increase the substrate doping there, as is evident from (1.5.12). The substrate doping below the source and drain areas can still be kept low, thus keeping the junction capacitances small.

As is obvious from the above discussion, the design of an implanted device is not an easy problem.⁵⁻⁸ One generally ends up with an implant, the characteristics of which are a compromise due to the above considerations.[†]

An nMOS transistor with a p -type implant is shown in Fig. 5.2a. The corresponding three-terminal structure is shown in Fig. 5.2b; consistent with our practice in past chapters, we will be using the study of this simpler structure as an intermediate step toward developing a model for the complete transistor.

An ion implant is characterized by the "effective dose" (number of implanted ions within the semiconductor per unit area, typically 10^{11} to 10^{13} per cm^2) and by the average kinetic energy of the ions as they leave the ion implanter (5 to 300 keV), which determines how deep they enter the semiconductor. The distribution of ion concentration within the semiconductor is initially nearly gaussian with respect to depth; after diffusion caused by high-temperature fabrication steps, it assumes the general shape shown by $N_i(y)$ in Fig. 5.3a, where y is the depth measured from the SiO_2 interface and N_{AB} is the doping concentration of the unimplanted substrate. Also shown in Fig. 5.3a is the total effective substrate doping, $N_{AB} + N_i(y)$. Because of the complicated shape of this profile, a detailed analysis is difficult.⁹⁻⁵⁶ However, it has been found that practical and useful results can be obtained by approximating this profile as shown in Fig. 5.3b,¹⁶ where N_i and d_i are appropriate constants chosen to make the resulting models as accurate as possible. As a starting point, d_i is sometimes chosen at the point where $N_i(y) = N_{AB}$ in Fig. 5.3a, or at a depth equal to the

[†]Two implants are sometimes used, to allow some degree of independence in setting the threshold voltage and controlling the punchthrough effect.

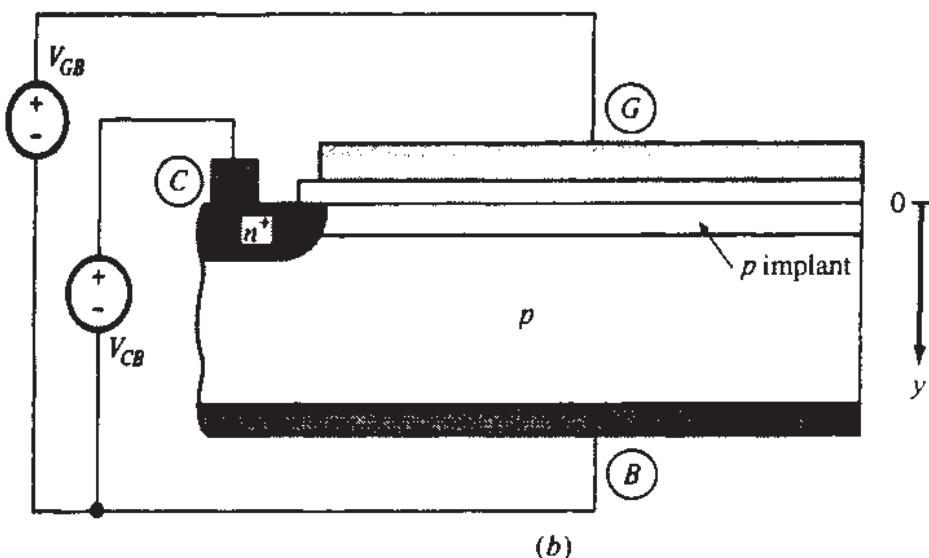
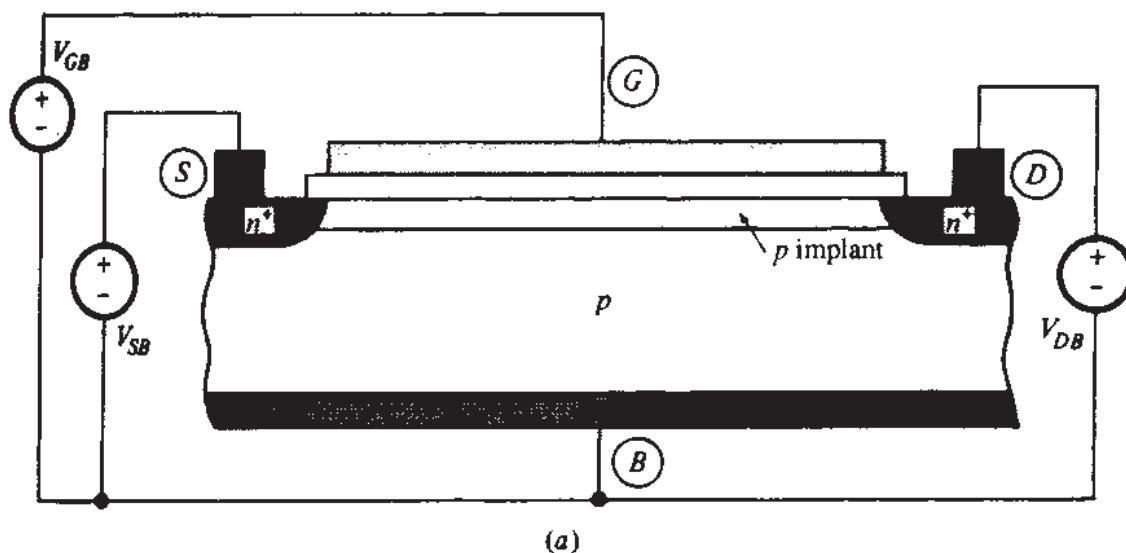


FIGURE 5.2

(a) An nMOS transistor with an ion-implanted channel; (b) three-terminal structure used to facilitate the study of (a).

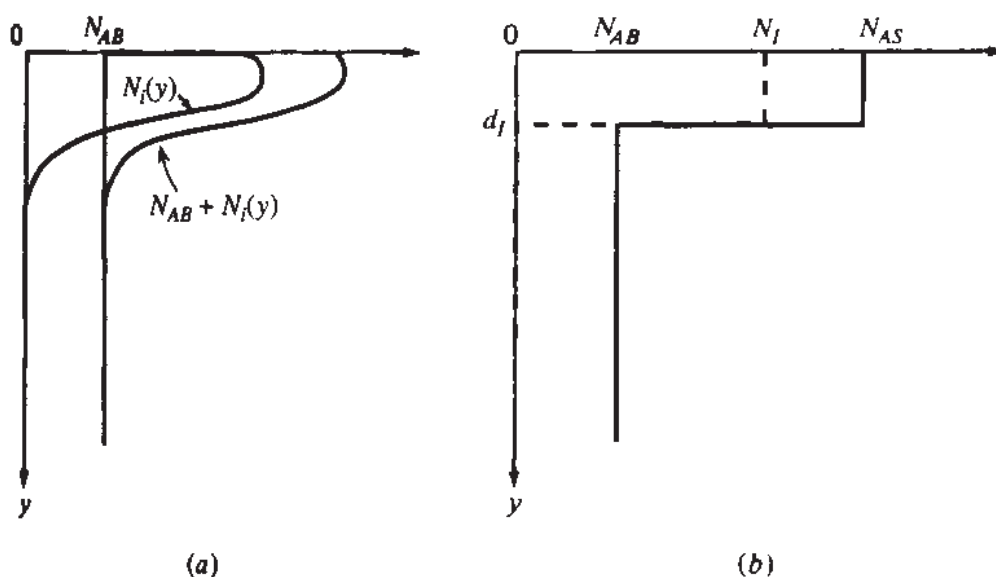


FIGURE 5.3

(a) Substrate doping concentration before implant, N_{AB} , implant concentration $N_I(y)$, and total doping concentration $N_{AB} + N_I(y)$ vs. depth from surface; (b) step approximation for the total concentration of (a).

mean plus the standard deviation of the distribution $N_i(y)$, assuming an initially gaussian shape; d_i can be, for example, $0.1 \mu\text{m}$. N_i is then chosen so that $N_i d_i$ is equal to the total effective dose [the integral under $N_i(y)$ in Fig. 5.3a]. We emphasize that these choices represent only a starting point. The best values for these parameters depend on what aspects of I - V characteristics need be modeled and the bias range over which reasonable accuracy is desired (see below).

5.2.2 Charges and Threshold Voltages

Consider first the corresponding three-terminal structure of Fig. 5.2b, assuming a very small V_{CB} . As V_{GB} is increased, a depletion region is formed at the surface. Its width increases with V_{GB} until a strong inversion layer is created, at which point the depletion region width becomes practically pinned to a certain value d_{Bm} . Assume that this value is smaller than the effective implant depth d_i , as shown in Fig. 5.4a.† If now V_{CB} is increased while strong inversion is maintained, the depletion region will widen and its bottom will eventually reach the bottom of the simplified implant profile. The critical value of V_{CB} at which this happens will be denoted by V_I and will be evaluated shortly. For $V_{CB} \leq V_I$, we can think of the gate oxide/implanted region as a device by itself, and consider the material below it as part of the substrate contact. The effective substrate doping of this device will be, from Fig. 5.3b,

$$N_{AS} = N_{AB} + N_I \quad (5.2.7)$$

We can use then the results of Chap. 3, with N_A replaced by N_{AS} . The width of the depletion region in strong inversion will be, from (3.4.8),

$$d_{Bm} = \sqrt{\frac{2\epsilon_s}{qN_{AS}}} \sqrt{\phi_{01} + V_{CB}}, \quad V_{CB} \leq V_I \quad (5.2.8)$$

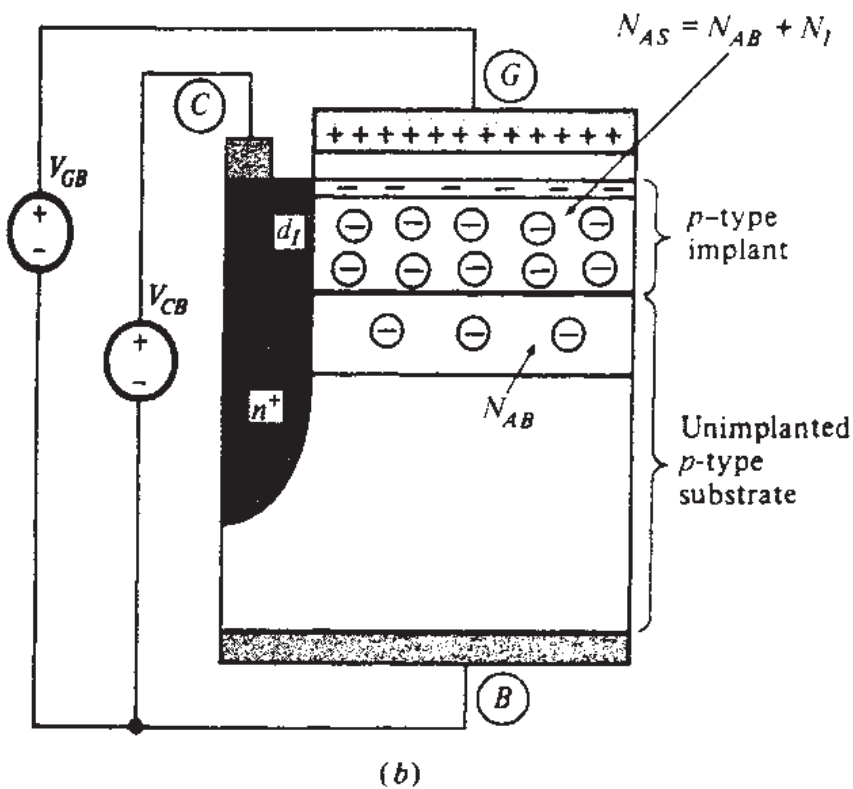
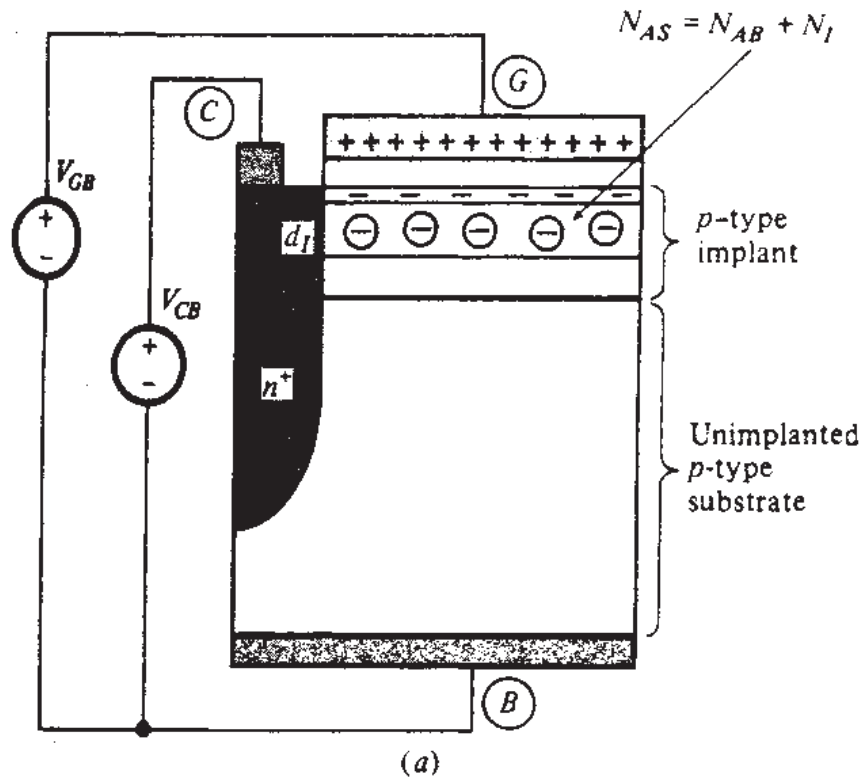
where ϕ_{01} will be somewhat larger than that corresponding to a substrate doping of N_{AB} (Sec. 2.5). The depletion region charge per unit area can be found as in Sec. 3.4.2:

$$Q'_B = Q'_{B1} = -\gamma_1 C'_{ox} \sqrt{\phi_{01} + V_{CB}}, \quad V_{CB} \leq V_I \quad (5.2.9)$$

where

$$\gamma_1 = \frac{\sqrt{2q\epsilon_s N_{AS}}}{C'_{ox}} \quad (5.2.10)$$

†Note that the depletion region around the n^+ region is not shown for simplicity.

**FIGURE 5.4**

The structure of Fig. 5.2b (*p*-type implant on a *p*-type substrate) in the strong-inversion region. (a) Small V_{CB} (depletion region totally within the implant); (b) large V_{CB} (depletion region extending outside the implant). The depletion region around the n^+ region is not shown for simplicity.

The inversion layer charge per unit area will be, again from Sec. 3.4.2,

$$Q'_I = -C'_{ox}(V_{GB} - V_{CB} - V_{T1}), \quad V_{CB} \leq V_I \quad (5.2.11)$$

where

$$V_{T1} = V_{FB1} + \phi_{01} - \frac{Q'_{B1}}{C'_{ox}} \quad (5.2.12a)$$

$$= V_{FB1} + \phi_{01} + \gamma_1 \sqrt{\phi_{01} + V_{CB}} \quad (5.2.12b)$$

with

$$\boxed{V_{FB1} \approx V_{FB}} \quad (5.2.13)$$

where V_{FB} is the "flat-band" voltage corresponding to the unimplanted substrate,† given by (5.2.4). The critical value V_I of V_{CB} at which the depletion region bottom reaches the bottom of the implant can be found by setting $V_{CB} = V_I$ and $d_{Bm} = d_I$ in (5.2.8), and solving for V_I . This gives

$$\boxed{V_I = \frac{qN_{AS}d_I^2}{2\epsilon_s} - \phi_{01}}^{\ddagger} \quad (5.2.14)$$

Increasing V_{CB} above V_I (assuming V_{GB} is large enough to maintain strong inversion) will move the depletion region bottom outside the implanted region, as shown in Fig. 5.4b. The results obtained above obviously do not apply in this case since they were obtained on the assumption that a single doping concentration characterizes the depletion region. In contrast, here we have a region with doping N_{AS} , followed by a region with doping N_{AB} . Let ψ_s be the surface potential, taken across *both* regions. A detailed analysis can be performed along the lines of Sec. 3.2; in fact, (3.2.2) to (3.2.4) hold unmodified. Such analysis leads to similar conclusions as in Sec. 3.2. In particular, in strong inversion we have again $\psi_s \approx \phi_0 + V_{CB}$, and we will use the value $\phi_0 = \phi_{01}$ as before for simplicity. The resulting *total* depletion region

†We are neglecting here the small contact potential between the implanted region and the unimplanted substrate. This is consistent with the overall level of approximations in the present analysis. We should note here that even when V_{CB} is such that the surface is neutral, one cannot expect that the electrostatic potential will be constant in the vertical direction *throughout* the substrate, because the dopant and mobile carrier concentration varies in that direction [see (1.2.11)]. Thus the corresponding energy bands (Appendix A) will not be "flat" throughout, in contrast to the case of uniform substrates (Sec. 2.2 and Appendix D). In implanted device work the name "flatband voltage" is a carryover from uniform substrate device discussions, and does not imply such flatness.

‡If the depletion region is within the implant for $V_{CB} = 0$ (as we have assumed in the beginning of this subsection), V_I will be positive. If, instead, the depletion region is deeper than the implant even for $V_{CB} = 0$, V_I will come out negative. This would indicate that the results obtained above for $V_{CB} \leq V_I$ are irrelevant for common applications (in which $V_{CB} \geq 0$), and we should instead consider the results for $V_{CB} > V_I$, which follow.

charge per unit area (as contributed by *both* regions in Fig. 5.4b) can be found using basic electrostatics (Appendix B), and is given by (Prob. 5.11)

$$Q'_B = Q'_{B2} = -qM - \gamma_2 C'_{ox} \sqrt{\phi_{01} - \frac{qMd_I}{2\epsilon_s} + V_{CB}}, \quad V_{CB} \geq V_I \quad (5.2.15)$$

where

$$\gamma_2 = \frac{\sqrt{2q\epsilon_s N_{AB}}}{C'_{ox}} \quad (5.2.16)$$

and

$$M = N_I d_I \quad (5.2.17)$$

is the implant "dose" corresponding to the rectangular shape of the distribution in Fig. 5.3b, measured in ions per cm² (or in cm⁻²). It is seen that (5.2.15) with $M = 0$ reduces to the corresponding equation for an unimplanted device (3.4.10).

One can find the inversion layer charge per unit area from (5.2.15), (3.2.2), (3.2.3), and (3.2.4) by eliminating among these equations the quantities Q'_G , ψ_{ox} , and Q'_B and using $\psi_s \approx \phi_{01} + V_{CB}$ as before. This gives

$$Q'_I = -C'_{ox}(V_{GB} - V_{CB} - V_{T2}), \quad V_{CB} \geq V_I \quad (5.2.18)$$

with

$$V_{T2} = V_{FB} + \phi_{01} - \frac{Q'_{B2}}{C'_{ox}} \quad (5.2.19a)$$

$$= V_{FB} + \phi_{01} + \frac{qM}{C'_{ox}} + \gamma_2 \sqrt{\phi_{01} - \frac{qMd_I}{2\epsilon_s} + V_{CB}} \quad (5.2.19b)$$

As a check, let us consider the case where d_I is negligibly small. Then (5.2.19b) reduces to (5.2.12b), with the only difference being that V_{FB} is augmented by the quantity qM/C'_{ox} . This makes sense, since $d_I = 0$ implies that all of the implanted ions are at the surface. Thus they act like the effective interface charge Q'_o in (5.2.4). Since these ions have a negative charge, their charge per unit area is $(-q)M$; thus V_{FB} is effectively augmented by qM/C'_{ox} .

It is seen that (5.2.18) is of the same form as (5.2.11). Also, (5.2.19b) has the same form of functional dependence on V_{CB} as (5.2.12b). Equation 5.2.19b can be written as follows:

$$V_{T2} = \left[V_{FB} + qM \left(\frac{1}{C'_{ox}} + \frac{d_I}{2\epsilon_s} \right) \right] + \left[\phi_{01} - \frac{qMd_I}{2\epsilon_s} \right] + \gamma_2 \sqrt{\left[\phi_{01} - \frac{qMd_I}{2\epsilon_s} \right] + V_{CB}} \quad (5.2.20)$$

In fact, consider a *fictitious* unimplanted device with a threshold voltage given by

$$V_{T2} = V_{FB2} + \phi_{02} + \gamma_2 \sqrt{\phi_{02} + V_{CB}} \quad (5.2.21)$$

Then it is easy to see that such a device will have a threshold behaving identically as (5.2.20) if we choose

$$\phi_{02} \equiv \phi_{01} - \frac{qMd_I}{2\epsilon_s} \quad (5.2.22)$$

and

$$V_{FB2} \equiv V_{FB} + qM \left(\frac{1}{C'_{ox}} + \frac{d_I}{2\epsilon_s} \right) \quad (5.2.23)$$

It is emphasized here that V_{FB2} is the flat-band voltage of the *fictitious* unimplanted device and should not be thought of as a flat-band voltage of the real one (similarly for the quantity ϕ_{02}). Nevertheless, we can take advantage of the convenient, familiar functional form of (5.2.21) and use it to describe the actual implanted device under consideration for $V_{CB} \geq V_I$, viewing ϕ_{02} and V_{FB2} as mere symbols defined by (5.2.22) and (5.2.23).† Note that

$$V_{FB2} > V_{FB1} \quad (5.2.24)$$

$$\phi_{02} < \phi_{01} \quad (5.2.25)$$

$$\gamma_2 < \gamma_1 \quad (5.2.26)$$

In conclusion, we can write Q'_I as follows:

$$Q'_I = -C'_{ox}[V_{GB} - V_{CB} - V_T(V_{CB})] \quad (5.2.27)$$

with

$$V_T(V_{CB}) = \begin{cases} V_{T1}(V_{CB}), & V_{CB} < V_I \\ V_{T2}(V_{CB}), & V_{CB} \geq V_I \end{cases} \quad (5.2.28a)$$

$$(5.2.28b)$$

†One should not, for that matter, be surprised if for some combination of process parameters ϕ_{02} is found to have a negative value. This simply means that the implant is such that $qMd_I/(2\epsilon_s) > \phi_{01}$, from (5.2.22). It is easy to check that the quantity under the square root in (5.2.21) never becomes negative in the domain of definition of that expression ($V_{CB} \geq V_I$).

where:

$$V_{Ti}(V_{CB}) = V_{FBi} + \phi_{0i} + \gamma_i \sqrt{\phi_{0i} + V_{CB}}, \quad i = 1, 2 \quad (5.2.29)$$

TRANSISTOR THRESHOLD. In a transistor (Fig. 5.2a), the gate-source threshold voltage can be found by using the fact that, at the source, the role of terminal C in Fig. 5.2b is played by the source terminal S; thus, replacing V_{CB} by V_{SB} in the above equations gives

$$V_T = \begin{cases} V_{T1}(V_{SB}), & V_{SB} < V_I \\ V_{T2}(V_{SB}), & V_{SB} \geq V_I \end{cases} \quad (5.2.30a)$$

$$(5.2.30b)$$

where

$$V_{Ti}(V_{SB}) = V_{FBi} + \phi_{0i} + \gamma_i \sqrt{\phi_{0i} + V_{SB}}, \quad i = 1, 2 \quad (5.2.31)$$

Since two regions are distinguished, and a number of approximations were made in deriving V_T for each region, it is important to check for possible anomalies at the critical point $V_{SB} = V_I$. A simple calculation shows that both $V_T(V_{SB})$ and dV_T/dV_{SB} are continuous at this point. Thus, no troublesome "kinks" will be caused in the transistor I - V characteristics by using this model for V_T (see below).

The quantity $V_T(V_{SB})$ is plotted in Fig. 5.5a. Two regions are clearly distinguishable.¹⁶ This behavior is verified by experiment. To obtain some intuition about

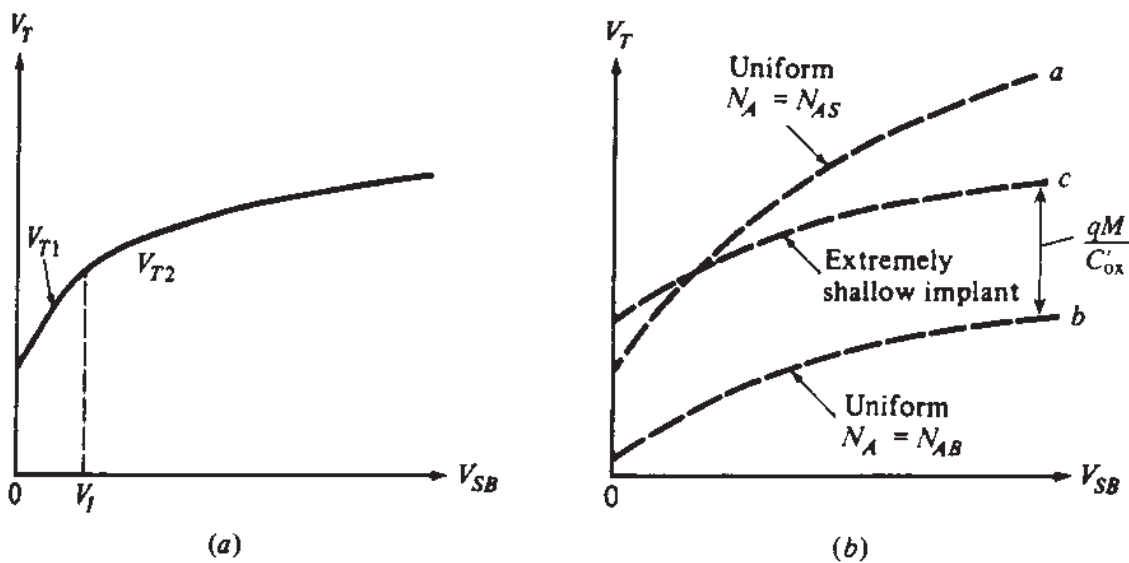


FIGURE 5.5

Threshold voltage vs. source-substrate bias for a p -type implant, p -substrate MOS transistor (Fig. 5.2a).

(a) For a step-approximated implant; (b) for three limit cases.

this behavior,¹⁶ consider an unimplanted device with uniform doping N_{AS} throughout the substrate. Then $V_T = V_{T1}$ for all V_{SB} values, resulting in curve *a* in Fig. 5.5*b*, characterized by γ_1 . Next, consider a device with uniform substrate concentration N_{AB} (no implant). Then $V_T(V_{SB})$ will look as shown by curve *b* in Fig. 5.5*b*, and will be characterized by a body effect coefficient γ_2 as given by (5.2.16). Finally, consider the substrate concentration unchanged from the value N_{AB} , and assume that the device has been subjected to an extremely shallow implant such that all ions end up practically at the oxide-semiconductor interface; the dose is assumed to be $M = N_I d_I$, the same as for Fig. 5.3*b*. Now, the total charge per unit area of these ions, $(-q)M$, simply serves to modify the value of the effective interface charge Q'_o , as already discussed. From (5.2.4) it follows that this will only result in a positive shift in V_{FB} by an amount qM/C'_{ox} . Hence, for this case, the plot of V_T vs. V_{SB} will be as shown by curve *c* in Fig. 5.5*b*. This curve results from a vertical shift of curve *b* by an amount equal to the change in V_{FB} . The substrate doping is still N_{AB} and, thus, the body effect coefficient is still γ_2 . The plot of V_T in Fig. 5.5*a* follows curve *a* for $V_{CB} < V_I$ and is close to curve *c* if V_{SB} is large. If d_I is reduced, the "transition point" in the plot will move to the left. For extremely shallow implants the curve in Fig. 5.5*a* will become practically the same as curve *c* in Fig. 5.5*b*. The device can then be characterized as discussed in previous chapters, with only a shift in V_{FB} . Comparing the curve in Fig. 5.5*a* to curve *b* in Fig. 5.5*b* (unimplanted device), it is seen that a *p*-type implant on a *p*-type substrate *increases* the extrapolated threshold voltage (for a given V_{SB}).†

The simple model we have presented is adequate for demonstrating the important threshold effects associated with the devices we are considering. (Other models are discussed elsewhere.²⁰) Of course, since a number of approximations were made in the course of deriving the model, one should allow for the adjustment of parameter values, notably M and d_I , to attain satisfactory agreement with measured results. A related discussion was presented in Sec. 4.15. Since the above model will be used in deriving *I-V* expressions, the final values of M and d_I one ends up with will depend on what aspects of the *I-V* characteristics must be modeled most accurately, over what range accuracy is desired, etc.‡ "Doping transformations" have been proposed for choosing the effective values of N_{AS} (or N_I) and d_I for good accuracy.^{27,34,42,51,54} In some of these techniques, these effective values are made functions of V_{SB} .

†This threshold-adjust implant is often in addition to a punchthrough control implant mentioned before. Since the latter would extend below the former, to depths where punchthrough would normally occur, the doping below the threshold-control implant considered in this chapter may not be N_{AB} as assumed, and, in fact, the picture of the depletion region there may be quite complicated. As a first-order correction, then, one may have to modify the value of N_{AB} appearing in our expressions. If the effective N_{AB} is allowed to be adjusted by a "parameter extraction" system (Chap. 10) for best matching to measurements, this modification would be done automatically.

‡One can be more specific if only particular aspects of the *I-V* characteristics need be modeled. An exact analysis of threshold voltages due to an arbitrary profile of any shape or depth in terms of dose and centroid parameters is given elsewhere.²⁵

Often, the expressions derived for V_T above, or similar ones, are used directly in the simplified strong-inversion model of Sec. 4.5.3. In many cases, reasonable accuracy can be obtained in this way, if parameter values are chosen by matching to measurements. It should be noted, though, that the derivation of the model of Sec. 4.5.3 was based on the assumption that the substrate is uniform, so, strictly speaking, that model is not valid in the case of implanted channels. The nonuniformity of the substrate can be taken into account in deriving a drain current model. For the interested reader, we now show how this can be done. Readers may choose to skip the next two subsections, depending on their modeling needs.

5.2.3 Drain-to-Source Current Model for Strong Inversion

Let us now consider the complete transistor in Fig. 5.2a. We assume operation in strong inversion. As was the case with the uniform-substrate device in Fig. 4.1a, if $V_{DB} \neq V_{SB}$, the effective reverse bias V_{CB} across the “field-induced junction” consisting of the inversion layer and the substrate will vary along the horizontal dimension. Thus the depletion region depth will vary, too. At any given point along this dimension, the depletion region may be shallower or deeper than the implant, depending on whether V_{CB} at that point is smaller or larger than V_T . Thus, Q'_I at that point will be given by (5.2.27), with V_T in it given by (5.2.28a) or (5.2.28b), correspondingly. We will take this into account in developing a model for the drain-source current.

NONSATURATION REGION. The drain current in nonsaturation can be determined from (4.5.7), which is repeated here:

$$I_{DSN} = \frac{W}{L} \int_{V_{SB}}^{V_{DB}} \mu(-Q'_I) dV_{CB} \quad (5.2.32)$$

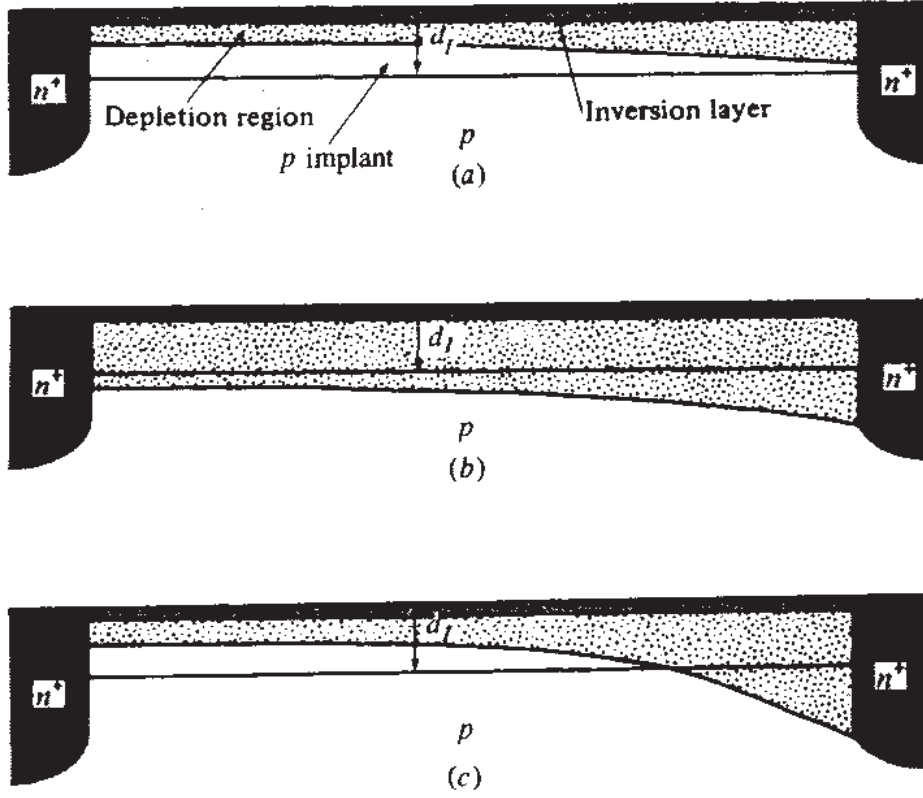
We will assume for now that μ is constant. [The mobility dependence on the gate field can be taken into account by using an effective mobility approach (Sec. 4.10).^{20,34,37,43,55}] With this assumption, and using Q'_I from (5.2.27), the above equation becomes

$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \int_{V_{SB}}^{V_{DB}} [V_{GB} - V_{CB} - V_T(V_{CB})] dV_{CB} \quad (5.2.33)$$

Assuming $V_{DB} > V_{SB}$ ($V_{DS} > 0$), we distinguish three cases:²⁰

1. $V_{SB} < V_{DB} \leq V_T$. In this case, neglecting edge effects very close to the source and the drain, the depletion region is inside the step-approximated implant throughout the length of the channel (Fig. 5.6a). Using (5.2.28a) and (5.2.12) in (5.2.33) gives

$$I_{DSN} = I_1(V_{SB}, V_{DB}) \quad (5.2.34)$$

**FIGURE 5.6**

Depletion region (shaded area) in an nMOS transistor with a p implant. The depletion region around the n^+ region is not shown for simplicity. (a) $V_{SB} < V_{DB} < V_I$; (b) $V_I < V_{SB} < V_{DB}$; (c) $V_{SB} < V_I < V_{DB}$.

where, for compactness in the formulation to follow, we use the notation

$$I_i(V_X, V_Y) = \frac{W}{L} \mu C'_{ox} \left\{ (V_{GB} - V_{Fbi} - \phi_{0i})(V_Y - V_X) - \frac{1}{2} (V_Y^2 - V_X^2) - \frac{2}{3} \gamma_i \left[(V_Y + \phi_{0i})^{3/2} - (V_X + \phi_{0i})^{3/2} \right] \right\} \quad (5.2.35)$$

This equation is in the form of (4.5.2b).

2. $V_I \leq V_{SB} < V_{DB}$. Here the depletion region edge is outside the step-approximated implant throughout the length of the channel (Fig. 5.6b). Using then (5.2.28b) and (5.2.21) in (5.2.33) we get

$$I_{DSN} = I_2(V_{SB}, V_{DB}) \quad (5.2.36)$$

where I_2 is given by (5.2.35) with $i = 2$.

3. $V_{SB} < V_I < V_{DB}$. In this case, the depletion region edge is inside the implant near the source and outside near the drain (Fig. 5.6c). It is at depth d_I at a point where

the reverse bias V_{CB} in the inversion layer is equal to V_I . Hence, we write (5.2.33) as follows:

$$I_{DSN} = \frac{W}{L} \mu C'_{ox} \left\{ \int_{V_{SB}}^{V_I} [V_{GB} - V_{CB} - V_{T1}(V_{CB})] dV_{CB} + \int_{V_I}^{V_{DB}} [V_{GB} - V_{CB} - V_{T2}(V_{CB})] dV_{CB} \right\} \quad (5.2.37)$$

Using (5.2.12) in the first integral and (5.2.21) in the second, we obtain

$$I_{DSN} = I_1(V_{SB}, V_I) + I_2(V_I, V_{DB}) \quad (5.2.38)$$

where I_i , $i = 1, 2$, is given by (5.2.35).

In summary, the nonsaturation current is given by:

$$I_{DSN} = \begin{cases} I_1(V_{SB}, V_{DB}), & V_{SB} < V_{DB} \leq V_I \\ I_2(V_{SB}, V_{DB}), & V_I \leq V_{SB} < V_{DB} \\ I_1(V_{SB}, V_I) + I_2(V_I, V_{DB}), & V_{SB} < V_I < V_{DB} \end{cases} \quad \begin{matrix} (5.2.39a) \\ (5.2.39b) \\ (5.2.39c) \end{matrix}$$

SATURATION REGION. Let us assume that V_{SB} is fixed, and that V_{DB} is raised. The onset of (forward) saturation can be determined as for the uniform-substrate transistor [see the development of (4.5.11)]. In other words, we consider that the inversion layer reaches "pinchoff" when the slope of I_{DN} with respect to V_{DB} reaches 0. If pinchoff occurs at a value of V_{DB} denoted by V_{P1} , such that $V_{P1} < V_I$, we have, equating the slope of (5.2.39a) to zero

$$V_{P1} = \left(-\frac{\gamma_1}{2} + \sqrt{\frac{\gamma_1^2}{4} + V_{GB} - V_{FB1}} \right)^2 - \phi_{01} \quad (5.2.40)$$

If, instead, pinchoff occurs at some value of V_{DB} larger than V_I , we find, by equating the slope of either (5.2.39b) or (5.2.39c) to zero,

$$V_{P2} = \left(-\frac{\gamma_2}{2} + \sqrt{\frac{\gamma_2^2}{4} + V_{GB} - V_{FB2}} \right)^2 \quad (5.2.41)$$

If $V_{SB} > V_I$, it is clear that (5.2.41) must be used in determining V_P , since we have assumed that $V_{DB} > V_{SB}$. If $V_{SB} < V_I$, (5.2.40) will be the equation to use if the value it predicts for a given V_{GB} is less than V_I . If, however, it predicts a value larger

than V_P , that value should be discarded and (5.2.41) should be used instead, as in that case pinchoff occurs while the depletion region edge near the drain end of the channel is outside the implant. If channel length modulation is neglected, the saturation current can be obtained by replacing the correct V_P value for V_{DB} in the appropriate equation among (5.2.39a) to (5.2.39c):

$$I'_{DS} = I_{DSN} \Big|_{V_{DB}=V_P} \quad (5.2.42)$$

and the complete model, covering both nonsaturation and saturation becomes

$$I_{DS} = \begin{cases} I_{DSN}, & V_{DB} \leq V_P \\ I'_{DS}, & V_{DB} > V_P \end{cases} \quad (5.2.43a)$$

$$(5.2.43b)$$

Reverse saturation (when V_{DB} is fixed and V_{SB} is raised) can be handled in a similar manner, with the role of V_{DB} played by V_{SB} (see Sec. 4.5).

The general shape of I_{DS} - V_{DS} characteristics is similar to that for unimplanted devices, but some peculiarities do arise. To make a simple discussion of these possible, we first introduce a simplified model.

5.2.4 Simplified Model for Strong Inversion

We can simplify the current equations and develop an approximate model related to that of (4.5.37). We can obtain such a model by using an approach similar to the one that led to (4.5.37) or even directly through an appropriate expansion of (5.2.35). However, here two different α parameters must be used: α_{I1} associated with the implanted region and α_{I2} associated with the unimplanted substrate. The details are outlined in the statement of Prob. 5.2. We find that in such a model we again obtain the nonsaturation current in the form of (5.2.39), only with $I_i(V_K, V_N)$ as follows:

$$I_i(V_X, V_Y) = \frac{W}{L} \mu C'_{ox} \left\{ [V_{GB} - V_X - V_{Ti}(V_X)](V_Y - V_X) - \frac{1}{2} \alpha_{Ii} (V_Y - V_X)^2 \right\} \quad (5.2.44)$$

where V_{Ti} , $i = 1, 2$, is given by (5.2.31). The parameters α_{I1} and α_{I2} must be carefully chosen, as the values of these parameters turn out to be crucial for guaranteeing continuity of the slope of the I_{DS} - V_{DB} characteristics as V_{DB} goes past the critical value V_P . Slope discontinuity, in addition to being physically incorrect, is undesirable because it leads to discontinuous small-signal parameters and can cause numerical problems when the model is incorporated in some computer programs for circuit analysis. This is discussed further in the statement of Prob. 5.2, where the following values are suggested for α_{I1} and α_{I2} :

$$\alpha_{I1} = 1 + \frac{\gamma_1}{\sqrt{\phi_{01} + V_I} + \sqrt{\phi_{02} + V_{SB}}} \quad (5.2.45a)$$

$$\alpha_{I2} = 1 + \frac{\gamma_2}{2\sqrt{\phi_{02} + V_I}} \quad (5.2.45b)$$

For the V_{SB} range in which (5.2.45a) is relevant ($V_{SB} < V_I$), it is easy to show that

$$\alpha_{I1} > \alpha_{I2} \quad (5.2.46)$$

which may be expected, since α_{I1} characterizes the heavily doped implanted region whereas α_{I2} characterizes the unimplanted substrate. Unlike what was done for uniform substrates [see (4.5.3) and associated discussion], here we are not completely free to choose the α parameters for best matching to experiment; our choices are limited by the requirements of continuity of I_{DS} and its derivatives. One cannot expect then that the above simple model will match measurements very well in all situations. Nevertheless, we will need the simplicity of the above model to illustrate some important aspects in the I - V characteristics of the devices we are considering. Interested readers may want to attempt an empirical improvement of this simple model.

In the saturation region we use (5.2.42) as before, with V_P determined as the value of V_{DB} at which the nonsaturation equation gives $dI_{DSN}/dV_{DB} = 0$. If $V_{SB} < V_I$ and pinchoff occurs at $V_P < V_I$, we have, working with (5.2.39a) and (5.2.44):

$$V_{P1} = V_{SB} + \frac{V_{GB} - V_{SB} - V_{T1}(V_{SB})}{\alpha_{I1}} \quad (5.2.47)$$

However, if the above equation predicts a V_P larger than V_I , we should realize that the pinchoff point is outside the implanted region, and thus we should have worked with (5.2.39c) and (5.2.44), obtaining

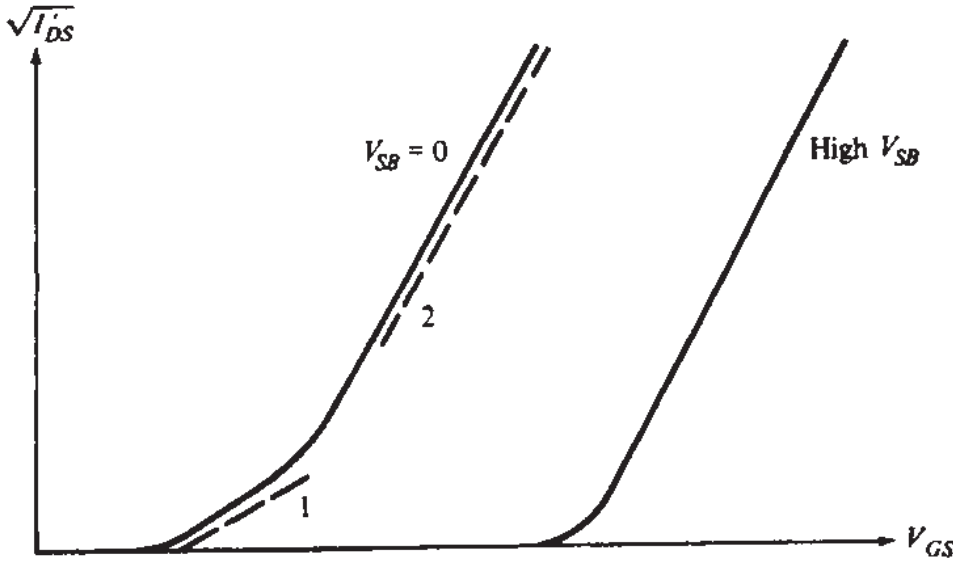
$$V_{P2} = V_I + \frac{V_{GB} - V_I - V_{T2}(V_I)}{\alpha_{I2}} \quad (5.2.48)$$

Finally, if $V_I < V_{SB} < V_{DB}$, we work with (5.2.39b) and (5.2.44) and obtain

$$V_{P3} = V_{SB} + \frac{V_{GB} - V_{SB} - V_{T2}(V_{SB})}{\alpha_{I2}} \quad (5.2.49)$$

The complete strong-inversion model is again given by (5.2.43), with the expressions for I_{DSN} and V_P appropriately chosen as explained.

For simplicity, sometimes the unimplanted-channel approximate model of Sec. 4.5.3 is used instead of the above models. In such a case, the parameters α_I , V_{FB} , ϕ_0 , and γ are assigned single "compromise" values. Such an approach can give satisfactory results if the dose is low. However, for high doses the results can be inaccurate. Shown in Fig. 5.7 is $\sqrt{I'_{DS}}$ in saturation, plotted vs. V_{GS} for a device with a *high* implant dose. Such behavior, which cannot be predicted by the model of Sec. 4.5.3, can be explained as follows. Assume first $V_{SB} = 0$. When V_{GB} is low, V_P is also low and

**FIGURE 5.7**

$\sqrt{I'_{DS}}$ as a function of V_{GS} in saturation for an nMOS transistor with a p implant, in the case of a heavy-dose deep implant, with V_{SB} as a parameter.

$V_P < V_I$. If I'_{DS} is the saturation current, using (5.2.42), (5.2.39a), (5.2.44), and (5.2.47), we find

$$\sqrt{I'_{DS}} = \sqrt{\frac{(W/L)\mu C'_{ox}}{2\alpha_{I1}}} [V_{GS} - V_{T1}(0)] \quad (5.2.50)$$

This is shown as broken curve 1 in Fig. 5.7. Consider now very large values of V_{GB} (with V_{SB} still zero). Then V_P will be much larger than V_I and will be predicted by (5.2.48). Using this equation in (5.2.42), with I_{DSN} from (5.2.39c) and (5.2.44), neglecting I_1 in comparison to I_2 , and using the fact that $V_{SB} = 0$, we find

$$\sqrt{I'_{DS}} \approx \sqrt{\frac{(W/L)\mu C'_{ox}}{2\alpha_{I2}}} \{V_{GS} - [V_I + V_{T2}(V_I)]\} \quad (5.2.51)$$

This is shown as broken curve 2 in Fig. 5.7. Note that the V_{GS} intercept in (5.2.51) is larger than that in (5.2.50), and the slope is larger since $\alpha_{I2} < \alpha_{I1}$, as mentioned above. A plot for a real device is close to curve 1 for low V_{GS} and close to curve 2 for large V_{GS} , as shown. Finally, when $V_{SB} > V_I$, (5.2.42), (5.2.39b), (5.2.44), and (5.2.49) give

$$\sqrt{I'_{DS}} = \sqrt{\frac{(W/L)\mu C'_{ox}}{2\alpha_{I2}}} [V_{GS} - V_{T2}(V_{SB})] \quad (5.2.52)$$

which is shown by the straight part of the right-hand solid curve in Fig. 5.7 (the bottom curved part is due to moderate and weak inversion). In this case, a single slope is predicted (assuming, of course, a mobility independent of V_{GS}).

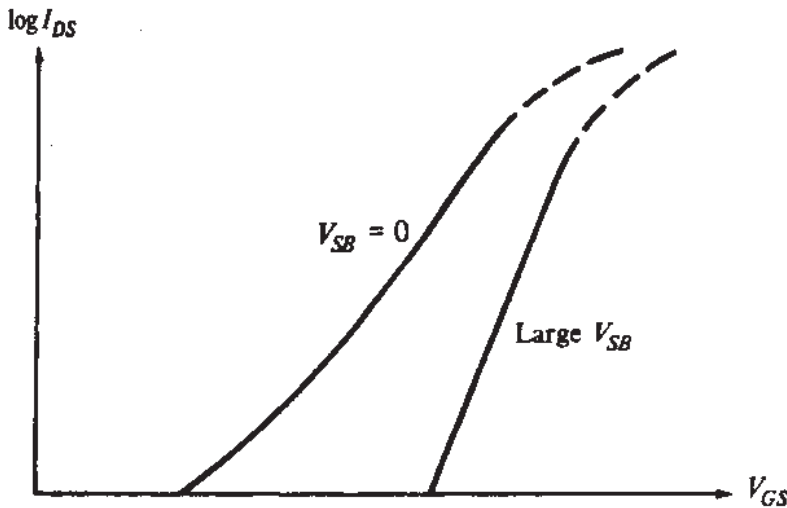


FIGURE 5.8
Weak inversion $\log I_{DS}$ vs. V_{GS} in saturation, with V_{SB} as a parameter, for an nMOS transistor with a p implant.

5.2.5 Weak Inversion

Let us now briefly look at the weak inversion characteristics. Depending on implant details, plots of $\log I_{DS}$ vs. V_{GS} in saturation may look as shown in Fig. 5.8. For large V_{SB} , the depletion region edge is outside the implant and the device behaves qualitatively as an unimplanted device. The slope of the curve is proportional to $1/n$ as before, with n given by (4.6.16), where $\gamma = \gamma_2$ as given by (5.2.16). However, for low V_{SB} the depletion region edge is inside the implant, $\gamma = \gamma_1$ as given by (5.2.10), and thus n and S (Sec. 4.6) are large and the slope is small. For low-voltage digital applications this can create problems since it now takes a larger “swing” of V_{GS} to change the current, and it may now be impossible to turn the device off. Note, also, that as V_{GS} increases the depletion region edge can be moving over a region with a widely varying concentration (from the “center” of the implant toward its “sides”). Then no simple value for n can be defined, and the slope of $\log I_{DS}$ with V_{GS} changes as shown; I_{DS} is far from being exponential with V_{GS} . This can create problems in some analog circuits, which rely on the near-exponential behavior normally expected of weak inversion.

From the above it should be clear that the threshold-adjust implant should be designed in such a way that the edge of the depletion region stays within the unimplanted region of the substrate. On the other hand, it will be seen in Chap. 6 that the depletion region extent should not be too large, since then short-channel effects can become pronounced. As already mentioned, a separate implant is often used to keep the depletion region horizontal extent small, at depths slightly less than the source-drain junction depths. As a compromise, then, the device can be designed so that the depletion region edge lies in between the threshold-adjust and the punchthrough-control implants.

5.3 DEPLETION nMOS TRANSISTORS

5.3.1 The Need for an n -Type Implant

In the preceding section we saw how, using a p -type implant with higher dopant concentration than the substrate, the threshold voltage can be increased from its value without implantation (e.g., -0.1 V) to a positive value that makes the transistor an

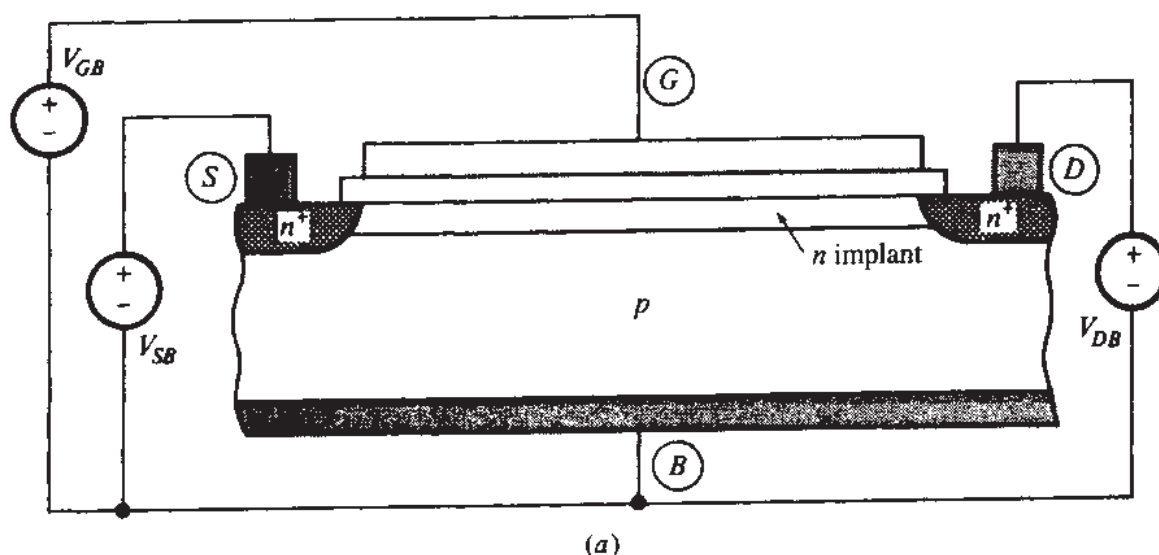


FIGURE 5.9

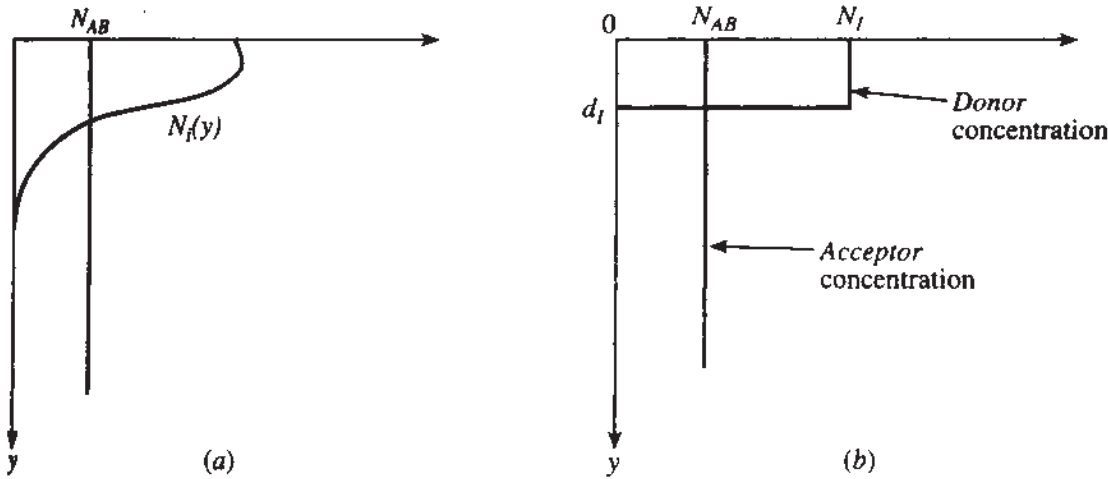
An nMOS transistor with an n -type implant.

enhancement-mode device. If it is desired to *decrease* V_{T0} instead (e.g., to make a depletion-mode device with a sufficiently *negative* V_{T0} †), the doping of the channel region must be changed in the opposite direction. This can be done by compensating the channel region by using donor ions, each of which cancels the effect of one ionized acceptor atom in the substrate (recall that the charge of an ionized acceptor atom is negative, whereas that of an ionized donor atom is positive). In fact, for a sufficiently negative threshold (e.g., -0.5 or -1 V), using numerical values in (5.2.1) indicates that Q'_B must be made positive, which means that a net concentration of donor atoms must exist in the channel. This leads to the device in Fig. 5.9.^{59,60} As seen, the n -type implant forms a conductive bridge between the n -type source and the n -type drain. In this device, a mode of operation which we have not yet encountered can occur: if the gate is made sufficiently negative, some of the electrons near the surface are repelled, and the surface becomes depleted. The conducting channel is thus limited to a path in the n implant *below* the surface. Devices in which this is the main operating region are called *buried channel* devices. By contrast, the transistor in Fig. 5.2a is a surface-channel device.

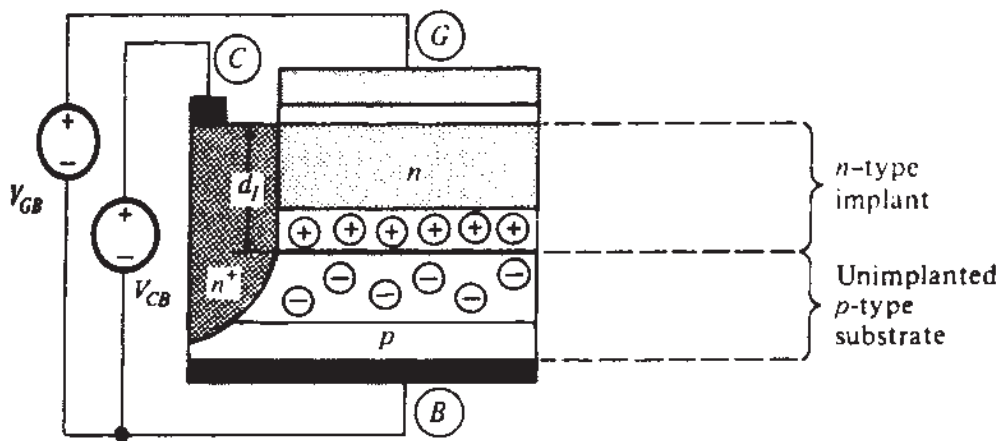
In the hypothetical case where all implanted ions (donors) end up at the surface, they simply change the value of the effective interface charge per unit area, Q'_o in (5.2.4), by the value $(+q)M$, where M is the effective dose; hence, V_{T0} shifts *down* by qM/C'_{ox} . More realistically, though, the implant profile will look as in Fig. 5.10a. It will be approximated as in Fig. 5.10b [here $N(y)$ and N_I represent *donor* concentrations]. With $N_I > N_{AB}$, the region between the surface and $y = d_I$ can thus be viewed as a uniformly doped n -type region, with an effective *donor* concentration of

$$N_{DS} = N_I - N_{AB} \quad (5.3.1)$$

†Such devices are used in some fancy CMOS technologies; they provide significant advantages in analog circuit design. They used to be standard in the older “nMOS” technologies.

**FIGURE 5.10**

(a) Substrate acceptor doping concentration before implant (N_{AB}) and donor implant concentration $N_I(y)$ vs. depth from surface; (b) step approximation for the donor implant concentration of (a).

**FIGURE 5.11**

A three-terminal structure corresponding to the nMOS, n -implant transistor of Fig. 5.9. The depletion region within the n^+ region is not shown for simplicity.

We will now analyze devices with the above type of implant.^{59-96,20,21,45,49,51} The modeling of such devices is rather complicated, so we must ask for the reader's patience.

5.3.2 Charges and Threshold Voltage

As usual, let us first consider a three-terminal structure, as shown in Fig. 5.11. For simplicity in drawing this figure, we have assumed that the interface charge Q'_o is zero. The p substrate and the n implant form a pn junction. The depletion region associated with this junction is shown in the figure. However, note that the depletion region around the n^+ region boundary is not shown, to keep the figure simple. The n

implant can be contacted through the n^+ region, which is contacted through terminal C. Thus, an external non-negative voltage V_{CB} acts as a reverse bias for the implant-substrate junction. The value of V_{CB} determines the width of the depletion region, and the magnitude of the charge per unit area on each side of the junction. To find the latter, we will use the two-sided step junction results from Appendix C, since N_{DS} is not necessarily much larger than N_{AB} . We easily obtain

$$Q'_J = \sqrt{2q\epsilon_s} \sqrt{\frac{N_{DS}N_{AB}}{N_{DS} + N_{AB}}} \sqrt{\phi_{bi} + V_{CB}} \quad (5.3.2)$$

where ϕ_{bi} is the built-in potential of the junction. Note that the charge per unit area is $+Q'_J$ at the top side of the junction and $-Q'_J$ at the bottom; hence, the total charge in the depletion region is zero.

The nondepleted part of the n implant contains mobile electrons. In the complete transistor, all these would be available for conduction, provided that no other part of the implant were depleted. However, if V_{GB} is made sufficiently negative, the resulting negative charges on the gate will repel some of the mobile electrons, and a depletion region (containing ionized donor atoms) will form at the top of the implant. If V_{GB} is made sufficiently negative, then all of the implant will be depleted (will be "pinched off"), and there will be no mobile electrons left in it. The value of V_{GB} at which this is attained is defined as the gate-substrate threshold voltage† V_{TB} , and the resulting situation for $V_{GB} \leq V_{TB}$ is shown in Fig. 5.12a.

According to this simple picture, if the n implant were part of a complete transistor, at $V_{GB} = V_{TB}$ there would be no mobile electrons available for conduction. This simple point of view is analogous to considering only strong inversion in unimplanted devices and saying that the current reduces to zero when the gate voltage is equal to the threshold voltage. While this view is not precise, it is often used to describe first-order behavior. It is to be understood that, if very low current operation is of interest, this view should be refined.‡

Let us now consider what happens at other V_{GB} values. As V_{GB} is raised above V_{TB} , the charges on the gate become less negative, and only part of the implant needs to be depleted in order to balance them. This is shown in Fig. 5.12b. If the structure were part of a transistor, the mobile electrons in the undepleted part of the implant would contribute to conduction. This undepleted part of the implant would thus form the channel of the transistor. As already mentioned, since this channel is below the surface, it is referred to as a *buried channel*, and devices operating in this mode are called *buried channel devices*.§

†In some of the literature, this quantity is also referred to as the "pinchoff voltage" in this context; this term should not be confused with the pinchoff voltage V_p in this book.

‡Among other things, the assumption that depletion regions edges are "sharp" would have to be reconsidered.

§If a low-dose implant is used, the upper part of the implant-substrate junction depletion region may reach all the way to the surface, even with $V_{CB} = 0$. In that case, increasing V_{GB} would eventually attract electrons to the surface, and one would have a surface-channel device. We are not considering this case here.

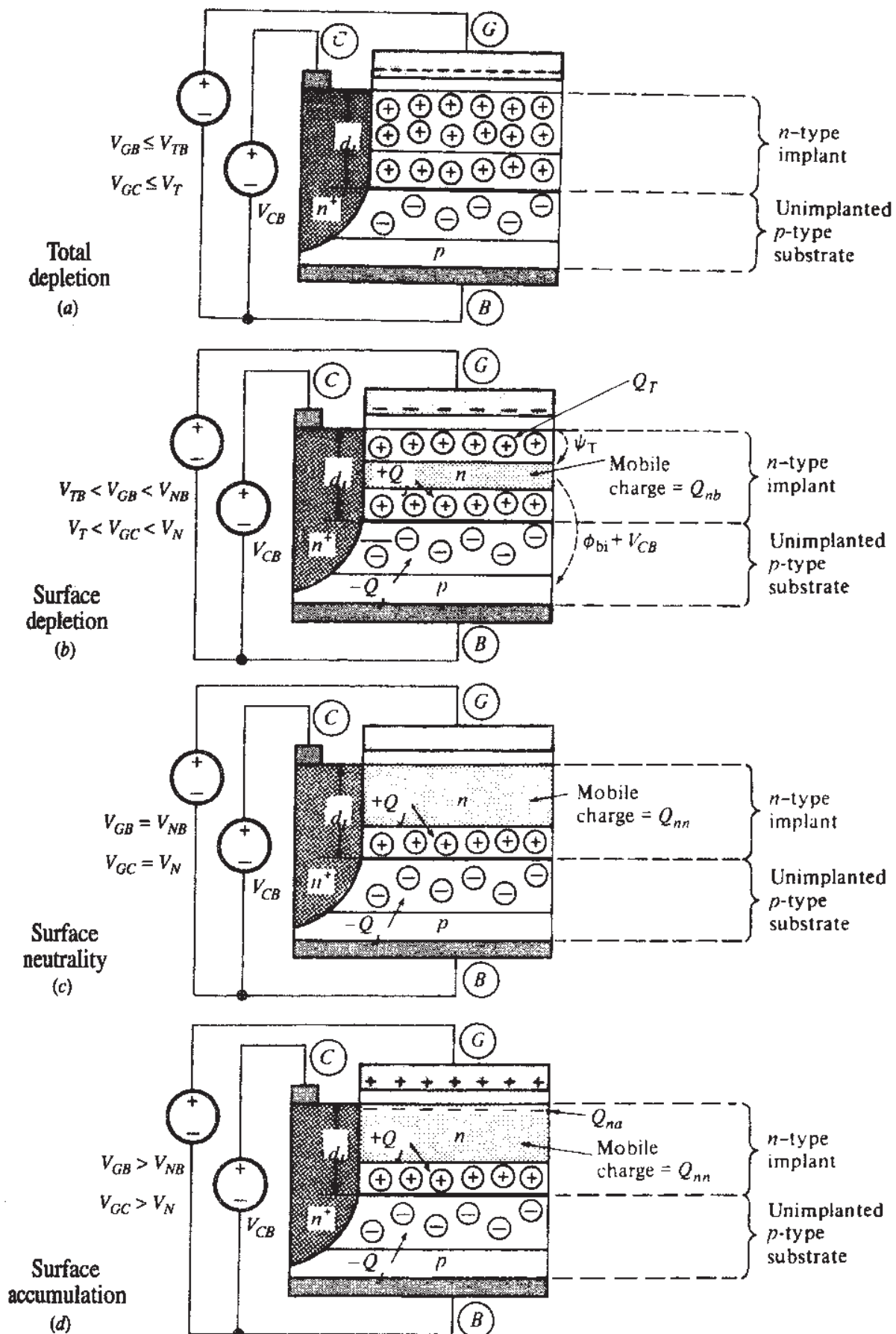


FIGURE 5.12

Condition below the oxide of an n -implant, p -substrate device for various gate-substrate voltage values. The depletion region around the n^+ region boundary is not shown, and Q'_o is assumed 0 for simplicity. (a) total depletion; (b) surface depletion; (c) surface neutrality; (d) surface accumulation.

If V_{GB} is raised further, a point will be reached at which no gate charges will exist (or, more properly, if the interface charge Q'_o is not assumed zero as in the figure, the gate charges will just balance Q'_o). Then the top of the depletion region will not be depleted, as shown in Fig. 5.12c. The value of V_{GB} at which this situation occurs is denoted by V_{NB} in the figure. Note that the undepleted part of the implant is neutral, as each mobile electron in it is balanced by one ionized donor atom.

If V_{GB} is raised still further, positive charges will be placed on the gate (or, if Q'_o is not assumed zero, $Q'_G + Q'_o$ will be positive). These charges must now be balanced by negative charges. Electrons will thus accumulate at the top of the implant, as shown in Fig. 5.12d. If the structure were part of a transistor, in this case the electrons in the accumulation layer would contribute to conduction, in addition to the mobile electrons in the neutral part of the implant.

We now develop the basic relations between the charges and the potentials of the three-terminal structure. The gate charge is related to the oxide potential as in (3.2.4):

$$Q'_G = C'_{ox}\psi_{ox} \quad (5.3.3)$$

To write the charge balance equation, we note that the charges in the substrate-implant depletion region balance each other out, and therefore do not contribute a net charge to the overall charge balance equation. Thus, the latter equation will involve only the gate charge Q'_G , the interface charge Q'_o , and any charge that might be present at the top of the implant region. In order to write the charge balance and potential balance equations, it is convenient to distinguish three cases:^{64,20,21}

SURFACE DEPLETION. In Fig. 5.12b, the charge at the top of the implant, which will be denoted by Q'_T , is due to ionized donor atoms and is positive. This is opposite from the depletion condition we have encountered in p -type substrates, and the potential ψ_T indicated in the figure has a *negative* sign. We can use a relation like (3.2.5a), but with appropriate sign changes:

$$Q'_T = Q'_{T1} = +\sqrt{2q\epsilon_s N_{DS}} \sqrt{-\psi_T} \quad (5.3.4)$$

With Q'_T as above, the charge balance equation is

$$Q'_G + Q'_o + Q'_T = 0 \quad (5.3.5)$$

The undepleted implant region in Fig. 5.12b does not contribute to the charge balance equation, since it is neutral: ionized impurity atoms in it are "covered" by mobile carriers of opposite charge.

The potential balance equation is obtained by going from the gate terminal through the device to the substrate terminal:

$$V_{GB} = \psi_{ox} + \psi_T + (\phi_{bi} + V_{CB}) + \phi_{MS} \quad (5.3.6)$$

where $\phi_{bi} + V_{CB}$ is the total potential drop across the depletion regions of the implant-substrate junction, as shown in Fig. 5.12b [see (1.5.11)], and ϕ_{MS} is the work function

potential difference corresponding to the gate and the *unimplanted* p -substrate materials, as always.

Using (5.3.3) to (5.3.6) and solving for Q'_T we obtain

$$Q'_T = -\frac{q\epsilon_s N_{DS}}{C'_{ox}} + \sqrt{\left(\frac{q\epsilon_s N_{DS}}{C'_{ox}}\right)^2 - 2q\epsilon_s N_{DS}(V_{GB} - V_{FB} - \phi_{bi} - V_{CB})} \quad (5.3.7)$$

where V_{FB} is given by (5.2.4). We now calculate the value of the gate-substrate threshold voltage V_{TB} . This voltage was defined as the value of V_{GB} which causes the two depletion regions in the n implant to meet in Fig. 5.12*b*, resulting in the situation of Fig. 5.12*a*. In this case, the total charge per unit area in these two depletion regions is equal to the total donor charge per unit area. The latter is $qN_{DS}d_I$, with N_{DS} given by (5.3.1). Thus, we have

$$Q'_T|_{V_{GB}=V_{TB}} + Q'_J = qN_{DS}d_I \quad (5.3.8)$$

Using in this expression (5.3.7) and (5.3.2), we can solve for V_{TB} . The expression obtained is of the form

$$V_{TB} = V_{CB} + V_T(V_{CB}) \quad (5.3.9)$$

A detailed expression for $V_T(V_{CB})$ is the subject of Prob. 5.5. That expression, if we assume $N_I \gg N_{AB}$, simplifies to the following²¹ (Prob. 5.5):

$$V_T(V_{CB}) = V_{T0} + \gamma_I(\sqrt{\phi_{bi} + V_{CB}} - \sqrt{\phi_{bi}}), \quad N_I \gg N_{AB} \quad (5.3.10)$$

where

$$V_{T0} = V_{FB} + \phi_{bi} - \frac{qN_{DS}d_I}{C'_{ox}} \left(1 + \frac{d_I C'_{ox}}{2\epsilon_s}\right) + \gamma_I \sqrt{\phi_{bi}} \quad (5.3.11)$$

and

$$\gamma_I = \left(1 + \frac{d_I C'_{ox}}{\epsilon_s}\right) \gamma \quad (5.3.12)$$

with γ being the body effect coefficient of an *unimplanted* device on the same p -type substrate. It is thus seen from (5.3.10) that the threshold voltage V_T increases with V_{CB} in a way similar to the threshold voltage of unimplanted devices. However, the effective body effect coefficient γ_I is larger than γ . This happens because the buried channel is farther away from the gate, in comparison to a surface channel device; thus, the gate releases some of its control to the substrate.

Note that the "body effect" is present in this device. When V_{CB} is increased, so is V_T . If, starting from Fig. 5.12b, V_{CB} is increased while V_{GB} is maintained constant, the channel will eventually disappear [i.e., V_{TB} in (5.3.9) will become equal to the applied V_{GB}]. Physically, this happens for two reasons: (1) the increasing V_{CB} widens the depletion region of the implant-substrate junction and (2) the increasing V_{CB} also widens the surface depletion region, as can be deduced from (5.3.7).

In a transistor, the gate-source threshold voltage can be found by replacing the role of terminal C by the source terminal S . Thus, using V_{SB} in lieu of V_{CB} in (5.3.10), we obtain

$$V_T = V_{T0} + \gamma_I \left(\sqrt{\phi_{bi} + V_{SB}} - \sqrt{\phi_{bi}} \right), \quad N_I \gg N_{AB} \quad (5.3.13)$$

with V_{T0} and γ_I as given above.

In some contexts, modeling stops at this point. The resulting expression for V_T is used in the transistor equations for the unimplanted device (for example, those in Sec. 4.5.3), without much justification. For the interested reader, below we will show how a more careful model can be developed. This will also be useful in order to show why, in some cases, models for undepleted devices seem to do an adequate job for the implanted transistors we are considering.

The mobile charge per unit area in the surface depletion condition (Fig. 5.12b) is due to the free electrons in the undepleted part of the implant, and will be denoted by Q'_{nb} . If the depletion regions were absent, the number of mobile electrons would be equal to the effective number of donor atoms. Thus Q'_{nb} would be equal to $-qN_{DS}d_I$. However, now some donor atoms have already been depleted. Since the electron population is reduced by one electron for each depleted atom, the remaining mobile electron charge per unit area is

$$Q'_{nb} = -(qN_{DS}d_I - Q'_T - Q'_J) \quad (5.3.14)$$

Before leaving the discussion of surface depletion, we parenthetically consider a special, normally undesired case. For given N_I and V_{CB} , devices with larger d_I will require more negative ψ_T to achieve the pinchoff condition illustrated in Fig. 5.12a. If d_I is too large, then ψ_T must attain such values that, before pinchoff can be achieved, surface inversion will occur, as shown in Fig. 5.13. (Note that the implanted region is n type, so the inversion layer consists of *holes*.) Once this inversion becomes strong, the surface potential and the width of the top depletion region do not change appreciably with V_{GB} . They become pinned to some value, as in the case of unimplanted devices. It now becomes impossible to pinch off the device by making V_{GB} more negative. Pinchoff can be achieved only by increasing V_{CB} , in which case the *bottom* depletion region in the n implant will widen until it touches the pinned edge of the top depletion region. To avoid the above problems, so that the device can be pinched off even with $V_{CB} = 0$, the simultaneous use of high dose and high energy for the implant is avoided.

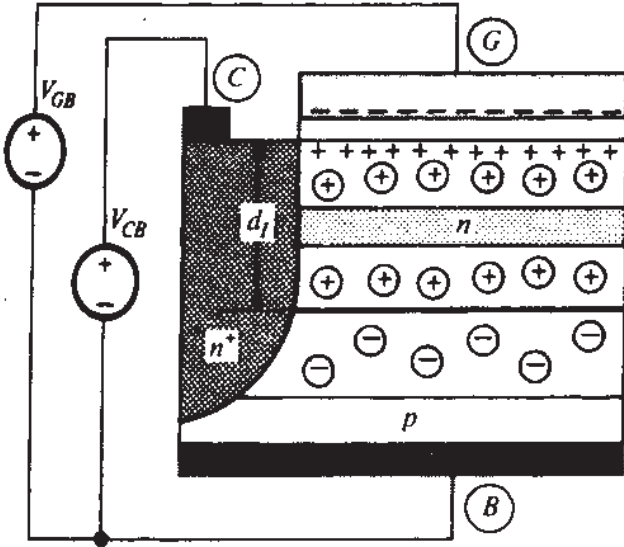


FIGURE 5.13

Illustration of the inability of the gate-substrate voltage to pinch off the channel because of a very high dose for an nMOS device with an n implant. The depletion region around the n^+ region is not shown for simplicity.

SURFACE NEUTRALITY. In Fig. 5.12c, the gate-substrate voltage is assumed to be at a value V_{NB} , such that the surface is neutral. Thus, the gate charge just balances the interface charge:

$$Q'_G + Q'_o = 0 \quad (5.3.15)$$

Since the surface is neutral, there is no voltage drop associated with it. Thus we can write

$$V_{GB} = \psi_{ox} + \phi_{bi} + V_{CB} + \phi_{MS} \quad (5.3.16)$$

Using (5.3.3) and (5.3.15) in this equation, we find that V_{GB} (the value of which will be denoted by V_{NB} for this case of surface neutrality) is given by

$$V_{NB} = \phi_{MS} - \frac{Q'_o}{C'_{ox}} + \phi_{bi} + V_{CB} \quad (5.3.17)$$

which becomes, using (5.2.4):

$$V_{NB} = V_N + V_{CB} \quad (5.3.18)$$

where

$$V_N = V_{FB} + \phi_{bi} \quad (5.3.19)$$

The mobile charge in the surface neutrality condition can be found as was done in order to obtain (5.3.14), only that now Q'_T has been reduced to zero. Denoting the mobile charge per unit area in this case by Q'_{nn} , we have:

$$Q'_{nn} = -(qN_{DS}d_I - Q'_J) \quad (5.3.20)$$

SURFACE ACCUMULATION.[†] With V_{GB} raised above V_{NB} (Fig. 5.12d), the charge at the top of the implant is due to mobile electrons in the accumulation layer. The thickness of this layer is very small and will be assumed to be infinitesimal; hence the potential needed across it to support it will be negligible (Appendix B). Thus (5.3.16) remains valid. The charge balance equation, though, must take into account the charge per unit area of the electrons in the accumulation layer; this charge will be denoted by Q'_{na} . Thus, we have

$$Q'_G + Q'_o + Q'_{na} = 0 \quad (5.3.21)$$

Using in this equation (5.3.3) and (5.3.16) and solving for Q'_{na} we obtain

$$Q'_{na} = -C'_{ox}(V_{GB} - V_{FB} - \phi_{bi} - V_{CB}) \quad (5.3.22)$$

This is only the mobile charge in the accumulation layer; in addition, there is the mobile charge in the rest of the undepleted implant, which has already been calculated for the surface neutrality condition above. That charge was found to be

$$Q'_{nn} = -(qN_{DS}d_I - Q'_I) \quad (5.3.23)$$

CONDITIONS IN TERMS OF V_{GB} AND V_{GC} . The relations that must be satisfied by V_{GB} and V_{GC} , for the various conditions discussed to be attained, are shown next to each part of Fig. 5.12d. The relations in terms of V_{GB} follow directly from our discussion. The relations in terms of V_{GC} follow easily from those, by observing that $V_{GC} = V_{GB} - V_{CB}$, and using (5.3.9) and (5.3.18).

5.3.3 Transistor Operation

We now consider a complete transistor as in Fig. 5.9, with $V_{GS} > V_T(V_{SB})$, so that a channel can exist next to the source (the various results obtained above can be applied to the source end of the channel, by replacing V_{CB} by V_{SB}). We assume $V_{DB} > V_{SB}$ ($V_{DS} > 0$). We will denote the effective reverse bias between a point in the neutral part of the channel and the unimplanted substrate by V_{CB} . We will thus be able to apply directly the formulas derived above. The value of V_{CB} increases from V_{SB} at the source to V_{DB} at the drain. Thus, the depletion region of the np junction will be deeper as we go toward the drain. Depending on the relative values of the terminal voltages, we can distinguish several regions of operation, which are discussed below.

Assume first that surface depletion exists next to the source. From Fig. 5.12b, this requires $V_T(V_{SB}) < V_{GS} < V_N$. Since V_{CB} increases toward the drain, $V_T(V_{CB})$ also increases, from (5.3.10), and tends to drive the drain end of the channel toward pinchoff. If the drain potential is not large enough to actually cause pinchoff, the channel will look as shown in Fig. 5.14a, where the shaded area represents the depletion region.²¹ If the drain potential exceeds a certain value, the channel will be

[†]Also referred to as *surface enhancement*.




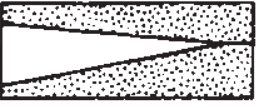

| | $V_T(V_{SB}) < V_{GS} < V_N$ Surface depletion | $V_{GS} > V_N$ |
|---------------|---|--|
| Nonsaturation |  $V_{DS} \leq V_{DS1}'$ (a) |  $V_{DS} < V_{GS} - V_N$ Surface accumulation (c) |
| | |  $V_{GS} - V_N < V_{DS} \leq V_{DS2}'$ Surface accumulation/depletion (d) |
| Saturation |  $V_{DS} > V_{DS1}'$ (b) |  $V_{DS} > V_{DS2}'$ Surface accumulation/depletion (e) |

FIGURE 5.14

Modes of operation for an nMOS depletion-mode transistor; shown is only the implant part. Depletion regions are shown shaded.

pinched off near the drain, as shown in Fig. 5.14b. The two regions of operation in Fig. 5.14a and b are called *nonsaturation* and *saturation*, respectively; the critical value of V_{DS} at the transition between them is denoted by V_{DS1}' . In saturation, the electrons are assumed to travel from the tip of the pinched-off channel through the depletion region to the drain, through a mechanism analogous to that for unimplanted devices in saturation. In analogy to the corresponding simplified picture for those devices, it can be argued here that, if the channel in Fig. 5.14b is long, the drain current in saturation will be practically constant and equal to the nonsaturation current obtained as V_{DS} approaches V_{DS1}' . A typical I_{DS} - V_{DS} characteristic for this case is shown by the lower curve in Fig. 5.15. A quantitative discussion of I_{DS} will follow later.

Assume now that V_{GB} is large enough to cause surface accumulation at the source. From Fig. 5.12d, this requires $V_{GS} > V_N$. If V_{DS} is small, surface accumulation can exist throughout the length of the channel, as shown in Fig. 5.14c. As the drain potential is raised, though, the depletion region widens and $|Q'_j|$ increases near the drain. From (5.3.22), with $V_{CB} = V_{DB}$, $|Q'_{na}|$ there will be decreasing and will eventually disappear. For V_{DB} values above this the drain end of the channel will exhibit surface depletion, as shown in Fig. 5.14d. For this to happen we must have, from Fig. 5.12b, $V_{GD} < V_N$. Writing $V_{GD} = V_{GS} - V_{DS}$, this gives the condition $V_{DS} > V_{GS} - V_N$.

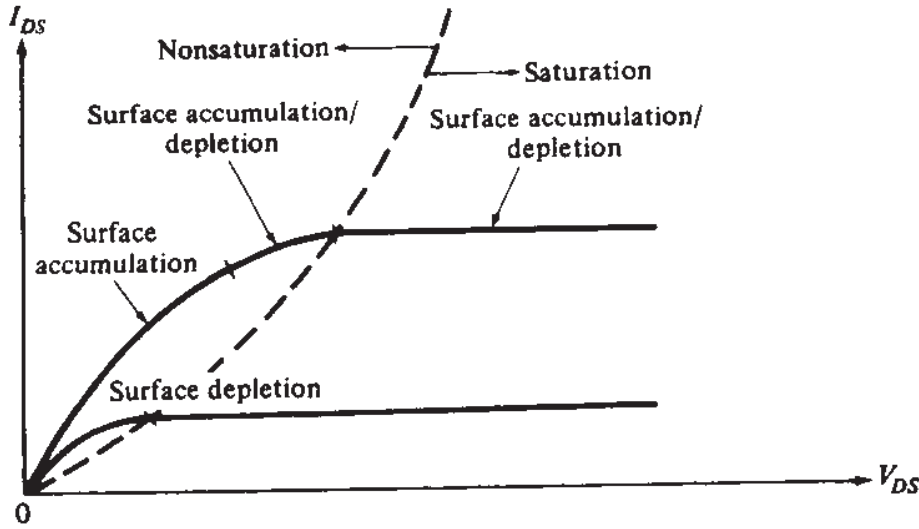


FIGURE 5.15

Relation of I - V characteristics to the modes of operation shown in Fig. 5.14.

Finally, further increase in the drain potential can increase $V_T(V_{DB})$ to the point where the channel becomes pinched off, as shown in Fig. 5.14e. The corresponding critical value of V_{DS} is denoted by V'_{DS2} . An I_{DS} - V_{DS} characteristic corresponding to the cases in Fig. 5.14c, d, and e is shown by the upper curve in Fig. 5.15.

We now show how the drain current can be calculated for the various regions summarized in Fig. 5.14.^{20,21} We first consider nonsaturation operation, corresponding to cases a, c, and d in Fig. 5.14. Here we cannot use (4.5.7), since there is no inversion layer involved. However, the role of Q'_i in that equation is now played by the total mobile charge per unit area, which, as we have already seen, can consist of either electrons flowing in the bulk of the implant, or of accumulation layer electrons flowing at its surface, or both. The mobilities involved will not be the same, since different scattering mechanisms are involved, as discussed in Sec. 4.10. Surface mobility is smaller than bulk mobility.

We now consider each of the three nonsaturation cases (see Figs. 5.14 and 5.15) separately.

1. *Surface depletion.*[†] In this case, shown in Fig. 5.14a, the total mobile charge per unit area is Q'_{nb} , given by (5.3.14). Thus instead of (4.5.7) we have

$$I_{DSN} = \frac{W}{L} \int_{V_{SB}}^{V_{DB}} \mu_B (-Q'_{nb}) dV_{CB} \quad (5.3.24)$$

where μ_B is the value of the mobility *in the bulk* of the n region.

[†]This mode of operation is sometimes also referred to as the *buried-channel mode*.

2. *Surface accumulation*, Fig. 5.14c. Here the total mobile charge consists of the surface charge Q'_{na} , given by (5.3.22), and the charge Q'_{nn} due to the mobile electrons in the undepleted part of the n -region bulk, given by (5.3.23). Hence, instead of (5.3.24) we have

$$I_{DSN} = \frac{W}{L} \int_{V_{SB}}^{V_{DB}} [\mu_s(-Q'_{na}) + \mu_B(-Q'_{nn})] dV_{CB} \quad (5.3.25)$$

where μ_s is the *surface* mobility.

3. *Surface accumulation/depletion*. This case is shown in Fig. 6.14d. Let V_{CBI} be the value of V_{CB} corresponding to that point in the channel where, as shown, we pass from accumulation to depletion behavior. At this point, $Q'_{na} = 0$ and from (5.3.22) we obtain

$$V_{CBI} = V_{GB} - V_{FB} - \phi_{bi} \quad (5.3.26)$$

To the left of this point, we have a total mobile charge per unit area of $Q'_{na} + Q'_{nn}$, from (5.3.22) and (5.3.23); to the right, the mobile charge per unit area is Q'_{nb} , from (5.3.14). Thus we have

$$I_{DSN} = \frac{W}{L} \int_{V_{SB}}^{V_{CBI}} [\mu_s(-Q'_{na}) + \mu_B(-Q'_{nn})] dV_{CB} + \frac{W}{L} \int_{V_{CBI}}^{V_{DB}} \mu_B(-Q'_{nb}) dV_{CB} \quad (5.3.27)$$

Using the expressions developed for Q'_{nb} , Q'_{na} , and Q'_{nn} in the above equations provides the nonsaturation current I_{DSN} . The value of V'_{DS} can be obtained as usual by setting $dI_{DSN}/dV_{DS} = 0$ or by setting the total mobile charge at the drain end of the channel equal to zero. The resulting formulation is complicated but, as before, the expressions can be simplified through appropriate series expansions. Simplified formulas thus obtained are shown in Table 5.1.²¹

As before, an effective *surface* mobility can be defined and made dependent on V_{GS} , as in Sec. 4.10.^{21,73,80†} The *bulk* mobility is often taken independent of V_{GS} .[‡]

For very small, fixed V_{DS} the plot of I_{DS} vs. V_{GS} , assuming constant mobility, is as shown by the broken line in Fig. 5.16a. For buried channels ($V_T \leq V_{GS} < V_N$) the device behaves as an unimplanted device qualitatively. However, although μ_B is larger than the surface mobility, the quantity $\mu_B C'_{ox}/(1 + \sigma)$, which characterizes surface depletion operation (see Table 5.1), is typically smaller than the corresponding quantity $\mu C'_{ox}$ of unimplanted devices. Intuitively, this happens because, as shown in Fig. 5.14a, the channel is further away from the gate than in the case of an

†A simple approximate model suggested for this²¹ is $\mu_{s,eff} = \mu_{s0}/[1 + \theta(V_{GS} - V_N)]$.

‡It has, however, been suggested⁸⁷ that, because the electrons move inside a thin slab of undepleted material (Fig. 5.14a), their mobility is limited to values below those for free bulk conduction. This is attributed to effects similar to those responsible for surface scattering, which become more severe as the "slab" becomes thinner. Since the thickness of the slab depends on V_{GS} , a dependence of μ_B on V_{GS} is suggested.

TABLE 5.1

Approximate strong-inversion model for n -implant- p -substrate devices^{21†}

$$V_T \leq V_{GS} < V_N:$$

$$\begin{aligned} I_{DS} &= \frac{W}{L} \frac{\mu_B C'_{ox}}{1 + \sigma} \left[(V_{GS} - V_T) V_{DS} - \frac{1}{2} \alpha V_{DS}^2 \right], & V_{DS} \leq V'_{DS1} \\ &= \frac{W}{L} \frac{\mu_B C'_{ox}}{1 + \sigma} \frac{(V_{GS} - V_T)^2}{2\alpha}, & V_{DS} \geq V'_{DS1} \\ V'_{DS1} &= \frac{V_{GS} - V_T}{\alpha} \end{aligned}$$

$$V_{GS} \geq V_N:$$

$$\begin{aligned} I_{DS} &= \frac{W}{L} \frac{\mu_B C'_{ox}}{1 + \sigma} \left\{ (V_{GS} - V_T) V_{DS} - \frac{1}{2} \alpha V_{DS}^2 + (r - 1) \left[(V_{GS} - V_N) V_{DS} - \frac{1}{2} V_{DS}^2 \right] \right\}, & V_{DS} < V_{GS} - V_N \\ &= \frac{W}{L} \frac{\mu_B C'_{ox}}{1 + \sigma} \left[(V_{GS} - V_T) V_{DS} - \frac{1}{2} \alpha V_{DS}^2 + \frac{1}{2} (r - 1) (V_{GS} - V_N)^2 \right], & V_{GS} - V_N \leq V_{DS} < V'_{DS2} \\ &= \frac{W}{L} \frac{\mu_B C'_{ox}}{1 + \sigma} \left[\frac{(V_{GS} - V_T)^2}{2\alpha} + \frac{1}{2} (r - 1) (V_{GS} - V_N)^2 \right], & V_{DS} \geq V'_{DS2} \\ V'_{DS2} &= \frac{V_{GS} - V_T}{\alpha} \end{aligned}$$

where

$$V_T = V_{T0} + \gamma_1 \left(\sqrt{\phi_{bi} + V_{SB}} - \sqrt{\phi_{bi}} \right)$$

$$\sigma = \frac{C'_{ox} d_I}{\epsilon_s} \left(\frac{C'_{ox} d_I}{2\epsilon_s} + 1 \right)$$

$$r = (1 + \sigma) \frac{\mu_S}{\mu_B}$$

$$\alpha = 1 + (1 + \sigma) \frac{\gamma}{4\sqrt{\phi_{bi}}}$$

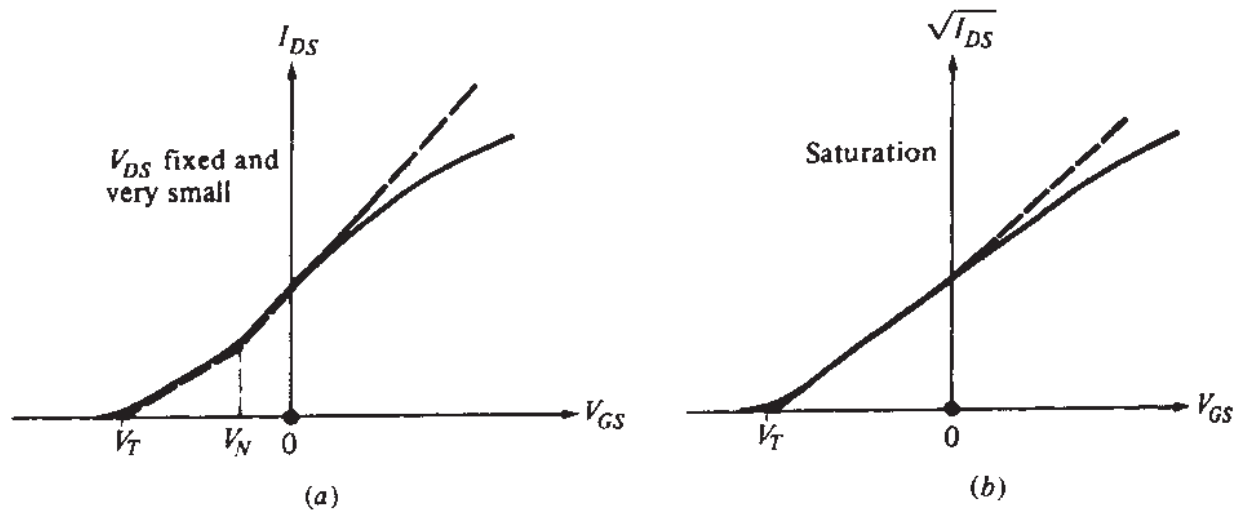
† $V_{DS} \geq 0$; channel length modulation not included.

FIGURE 5.16

Characteristics of an nMOS transistor with an n implant. (a) I_{DS} as a function of V_{GS} with very small V_{DS} ; (b) saturation $\sqrt{I_{DS}}$ vs. V_{GS} . Dashed line: model with constant mobility; solid line: real device.

unimplanted device, and thus the influence of the gate is somewhat reduced. Once surface accumulation occurs, however ($V_{GS} \geq V_N$), variations of the channel charge with V_{GS} occur at the surface, and, thus, the slope is determined by the surface mobility and the oxide thickness only. The solid line in Fig. 5.16a represents a real device, and no breakpoint is, of course, observed in this case. Mobility degradation at high V_{GS} can be seen. The deviation from straight line at very low V_{GS} is due to the presence of diffusion currents, which we have not included in the simple model we have presented. This phenomenon is reminiscent of weak inversion in unimplanted devices.⁸⁶

For the plot of $\sqrt{I_{DS}}$ vs. V_{GS} in saturation, the model predicts again a changing slope effect, although not so pronounced, as shown by the broken line in Fig. 5.16b. The slight upward trend of the slope tends to be counteracted by the degradation of mobility with V_{GS} , and the resulting behavior is as shown by the solid line, a significant part of which is practically straight. Thus, a saturation equation like (4.5.37b) can again be used in such cases in the form

$$I_{DS} = k(V_{GS} - V_T)^2 \quad (5.3.28)$$

only with the constant of proportionality k somewhat reduced from that of an unimplanted device, and V_T negative for low V_{SB} . For some depletion devices, the above equation has been shown to be accurate, even more so than for enhancement devices on the same substrate.⁷⁶ In fact, in some circuit analysis computer programs the complete equation (4.5.37), in both nonsaturation and saturation, is used to model implanted depletion devices, using a negative V_T . Depending on device details, however, significant errors can result from such approximations. Models involving 3/2 powers, analogous to (4.5.2), have also been proposed^{64,74} (Prob. 5.9).

It should be noted that the model of Table 5.1 is not valid for devices having the problem illustrated in Fig. 5.13. In such devices, a channel will exist when V_{SB} is small, even if V_{GS} is made very negative. The behavior shown in Fig. 5.17 will then be observed. Proper device design avoids such problems.

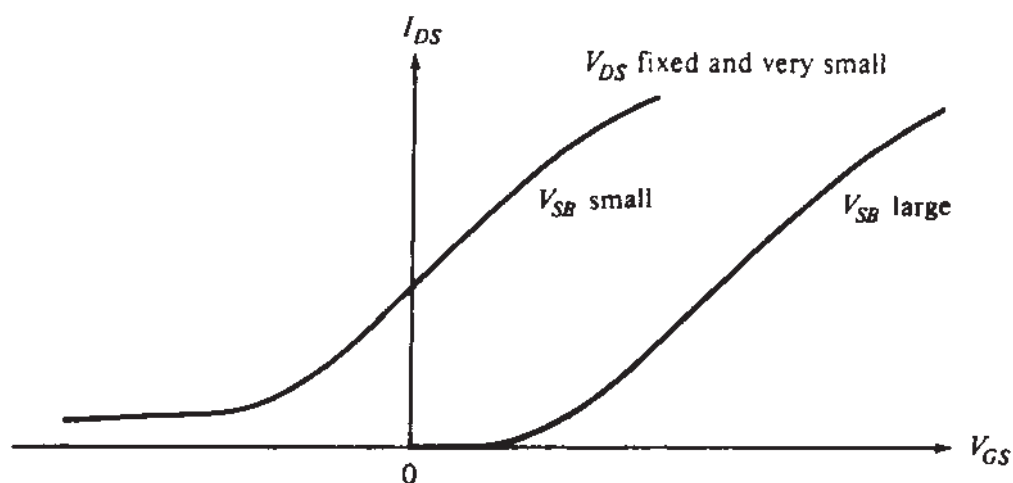


FIGURE 5.17

I_{DS} vs. V_{GS} , with V_{SB} as a parameter, for the n -implant, nMOS transistor with a very large implant dose (see also the corresponding three-terminal structure of Fig. 5.13).

The model we have presented is not valid when the two depletion regions are about to meet. This is both because then our assumptions of “abrupt” depletion region edges are inadequate, and because diffusion currents become important. Depending on the value of V_{SB} , the “meeting point” of the two regions can occur at various depths, a fact that affects both the distance from the gate and the effective implant concentration at that point (where most of the current flows). Thus, the dependence of I_{DS} on V_{SB} in this part of the operating range is complicated. Models valid in all regions of operation (analogous to the charge sheet model of Sec. 4.3.1) have been proposed for the depletion device.^{86,95}

The n -type implants on p substrates are usually used for making depletion-mode devices. However, it is possible to end up with enhancement-mode devices after lowering the threshold, if one starts with a sufficiently high V_{FB} .^{70,82,45,51}

5.4 ENHANCEMENT pMOS TRANSISTORS

5.4.1 Surface-Channel Enhancement-Mode pMOS

Figure 5.18a shows an unimplanted pMOS device with a p^+ poly gate. This is the opposite situation of that shown in Fig. 5.2a, and with lightly doped substrates it would

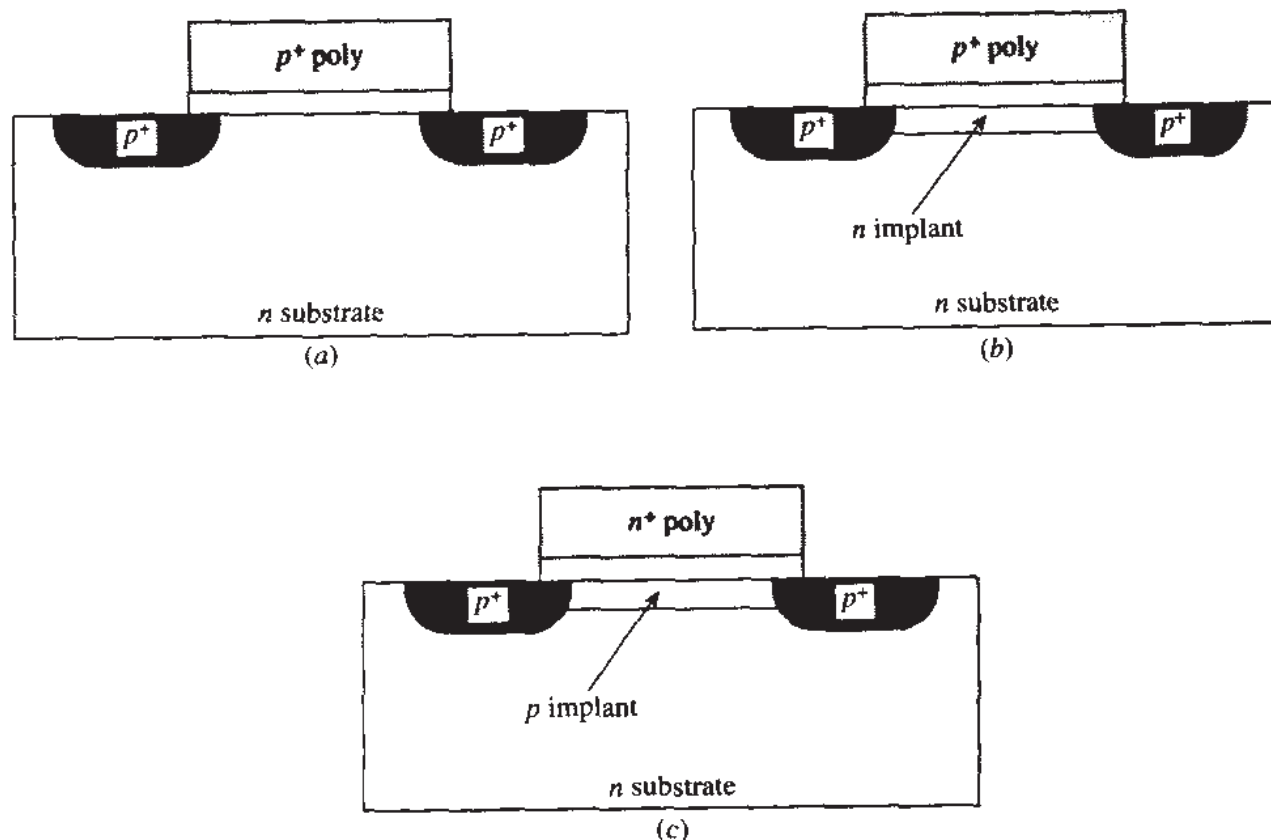


FIGURE 5.18

pMOS transistors. (a) with p^+ gate and unimplanted channel; (b) with p^+ gate and n implant; (c) with n^+ gate and p implant.

normally give a slightly positive V_{T0} . We saw in Sec. 5.2 that to make a proper enhancement nMOS device, we needed to use an implant of the same type as the substrate (which was p type for that device), only with increased doping concentration; similarly, then, to make a proper enhancement-mode pMOS device (with a V_{T0} of, say, -0.5 V), one should raise the n -type doping concentration in the channel, as shown in Fig. 5.18b. This results in a surface-channel device and is the way pMOS transistors are made in several modern fabrication processes. The modeling of such devices can be done as in Sec. 5.2, and the expressions there can be used with appropriate sign changes, in the spirit of Sec. 4.13.

5.4.2 Buried-Channel Enhancement-Mode pMOS

In some CMOS processes, the poly gates used for both nMOS and pMOS devices are doped n^+ . This makes ϕ_{MS} in (5.2.4) about 1.1 V more negative than for p^+ poly (see Secs. 2.2 and 1.4), and results in a very negative threshold voltage (e.g., -1.2 V). In this case, we need to increase the threshold (say, to a desired value of -0.5 V). Using values in (5.2.1) shows that Q'_B would have to be made negative. This suggests the use of a p -type implant as shown in Fig. 5.18c.^{61,82,45} Just like the device of Fig. 5.9, this is a buried channel device; it is often referred to as *compensated*. Such a transistor can suffer from significant short-channel effects (Chap. 6), so many technologies today use p^+ gates for pMOS devices, as discussed in the previous subsection.

Depletion-mode pMOS transistors are not used in common fabrication technologies.

REFERENCES

1. K. G. Aubuchon, *Proceedings of the International Conference on Properties and Use of M.I.S. Structures*, Grenoble, France, p. 575, 1969.
2. R. W. Bower and H. G. Dill, *International Electron Devices Meeting*, paper 16.6 (unpublished), 1966.
3. M. R. MacPherson, "The adjustment of MOS transistor threshold voltage by ion implantation," *Applied Physics Letters*, vol. 18, pp. 502–504, 1971.
4. T. Masuhara, M. Nagata, and N. Hashimo, "A high-performance n-channel MOS LSI using depletion type load elements," *IEEE Journal of Solid-State Circuits*, vol. SC-7, pp. 224–231, June 1972.
5. W. S. Ruska, *Microelectronics Fabrication*, McGraw-Hill, New York, 1987.
6. S. M. Sze, *VLSI Technology*, 2d ed., McGraw-Hill, New York, 1988.
7. J. Y. Chen, *CMOS Devices and Technology for VLSI*, Prentice-Hall, Englewood Cliffs, N.J., 1990.
8. S. K. Ghandhi, *VLSI Fabrication Principles—Silicon and Gallium Arsenide*, 2d ed., Wiley, New York, 1994.
9. M. R. MacPherson, "Threshold shift calculation for ion implanted MOS devices," *Solid-State Electronics*, vol. 15, pp. 1319–1326, 1972.
10. M. Kamoshida, "Threshold voltage and gain term β of ion implanted n-channel MOS transistors," *Applied Physics Letters*, vol. 22, pp. 404–405, 1973.
11. G. Doucet and F. Van de Wiele, "Threshold voltage of nonuniformly doped structures," *Solid-State Electronics*, vol. 16, pp. 417–423, 1973.
12. P. P. Peressini and W. S. Johnson, "Threshold adjustment of n-channel enhancement mode FETs by ion implantation," *Technical Digest*, International Electron Devices Meeting, Washington, D.C., pp. 467–486, 1973.

13. R. H. Dennard, F. H. Gaensslen, H. Yu, V. L. Rideout, E. Bassons, and A. R. LeBlanc, "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. SC-9, pp. 256-268, 1974.
14. E. C. Douglas and A. G. F. Dingwall, "Ion implantation for threshold control in COSMOS circuits," *IEEE Transactions on Electron Devices*, vol. ED-21, pp. 324-331, 1974.
15. M. Kamoshida, "Electrical characteristics of boron-implanted n-channel MOS transistors," *Solid-State Electronics*, vol. 17, pp. 621-626, 1974.
16. V. L. Rideout, F. H. Gaensslen, and A. LeBlanc, "Device design considerations for ion-implanted n-channel MOSFETs," *IBM Journal of Research and Development*, p. 50, January 1975.
17. G. Doucet, F. Van de Wiele, and P. Jespers, "Theoretical and experimental study of MOS transistors nonuniformly doped by SILOX technique," *Solid-State Electronics*, vol. 19, pp. 191-199, 1976.
18. R. R. Troutman, "Ion-implanted threshold tailoring for insulated gate field-effect transistors," *IEEE Transactions on Electron Devices*, ED-24, pp. 182-192, 1977.
19. H. Feltl, "Onset of heavy inversion in MOS devices doped nonuniformly near the surface," *IEEE Transactions on Electron Devices*, vol. ED-24, pp. 288-289, 1977.
20. E. Demoulin and F. Van de Wiele, "Ion implanted MOS transistors," in *Process and Device Modelling for Integrated Circuit Design*, F. Van de Wiele, W. L. Engl, and P. G. Jespers (editors), Noordhoff, Leyden, The Netherlands, 1977.
21. G. Merckel, "Ion implanted MOS transistors—depletion mode devices," in *Process and Device Modelling for Integrated Circuit Design*, F. Van de Wiele, W. L. Engl, and P. G. Jespers (editors), Noordhoff, Leyden, The Netherlands, pp. 617-676, 1977.
22. J. R. Brews, "Threshold shifts due to nonuniform doping profiles in surface channel MOSFET's," *IEEE Transactions on Electron Devices*, vol. ED-26, p. 1696, 1979.
23. M. Nishida and M. Aoyane, "An improved definition for the onset of heavy inversion in a MOS structure with nonuniformly doped semiconductors," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 1222-1230, 1980.
24. S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.
25. J. R. Brews, "Physics of the MOS transistor," chap. 1 in *Silicon Integrated Circuits, Part A*, D. Kahng (editor), Applied Solid-State Science Series, Academic Press, New York, 1981.
26. L. M. Dang and H. Iwai, "Modeling the impurity profile of an ion-implanted IGFET for the calculation of threshold voltages," *IEEE Transactions on Electron Devices*, vol. ED-28, pp. 116-117, January 1981.
27. P. K. Chatterjee, J. E. Leiss, and G. W. Taylor, "A dynamic average model for the body effect in ion implanted short-channel ($L = 1 \mu\text{m}$) MOSFET's," *IEEE Transactions on Electron Devices*, vol. ED-28, pp. 606-607, May 1981.
28. K-Y. Fu, "A new analysis of the threshold voltage for non-uniform ion-implanted MOSFET's," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 1810-1813, November 1982.
29. K. Shenai, "Analytical solutions for threshold voltage calculations in ion-implanted IGFET's," *Solid-State Electronics*, vol. 26, pp. 761-766, 1983.
30. A. H. Marshak and R. Shrivastava, "On threshold and flat-band voltages for MOS devices with polysilicon gate and nonuniformly doped substrate," *Solid-State Electronics*, vol. 26, pp. 361-364, 1983.
31. F. Van de Wiele, "On the flat-band voltage of MOS structures on nonuniformly doped substrate," *Solid-State Electronics*, vol. 27, pp. 824-826, 1984.
32. D. A. Divekar and R. I. Dowell, "A depletion-mode MOSFET model for circuit simulation," *IEEE Transactions on Computer-Aided Design*, vol. CAD-3, p. 80, January 1984.
33. D. A. Antoniadis, "Calculation of threshold voltage in nonuniformly doped MOSFET's," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 303-307, March 1984.
34. P. Ratnam and C. A. T. Salama, "A new approach to the modeling of nonuniformly doped short-channel MOSFET's," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 1289-1298, September 1984.
35. C. R. Viswanathan, B. C. Burkey, G. Lubberts, and T. J. Tredwell, "Threshold voltage in short channel MOS devices," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 932-940, 1985.
36. H. J. Park and C.-K. Kim, "An empirical model for the threshold voltage of enhancement NMOS-FET's," *IEEE Transactions on Computer-Aided Design*, vol. CAD-4, p. 629, October 1985.

37. C-Y. Wu and Y-W. Daih, "An accurate mobility model for the I - V characteristics of n-channel enhancement-mode MOSFETs with single-channel boron implantation," *Solid-State Electronics*, vol. 28, pp. 1271-1278, 1985.
38. D. M. Rogers, J. D. Hayden, and D. D. Rinerson, "Model for the channel-implanted enhancement-mode IGFET," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 955-964, July 1986.
39. G. T. Wright, "Physical and CAD models for the implanted-channel VLSI MOSFET," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 823-833, April 1987.
40. A. Dasgupta and S. K. Lahiri, "An analytical threshold voltage model of short-channel MOSFET's with implanted channels," *IEEE Transactions on Electron Devices*, vol. ED-34, p. 1177, May 1987.
41. R. V. Booth, M. H. White, H.-S. Wong, and T. J. Krutsick, "The effect of channel implants on MOS transistor characterization," *IEEE Transactions on Electron Devices*, vol. ED-34, p. 2501, December 1987.
42. N. D. Arora, "Semi-empirical model for the threshold voltage of a double implanted MOSFET and its temperature dependence," *Solid-State Electronics*, vol. 30, pp. 559-569, 1987.
43. T. J. Krutsick and M. H. White, "Consideration of doping profiles in MOSFET mobility modeling," *IEEE Transactions on Electron Devices*, vol. 35, pp. 1153-1155, July 1988.
44. S. Karmalkar and K. N. Bhat, "A process-parameter-based circuit simulation model for ion-implanted MOSFET's and MESFET's," *IEEE Journal of Solid-State Circuits*, vol. 24, pp. 139-145, February 1989.
45. H. C. de Graaff and F. M. Klaassen, *Compact Transistor Modelling for Circuit Design*, Springer-Verlag, Vienna, 1990.
46. P.-S. Lin and C.-Y. Wu, "A new simplified two-dimensional model for the threshold voltage of MOSFET's with nonuniformly doped substrate," *IEEE Transactions on Electron Devices*, vol. 38, p. 1376, June 1991.
47. S. Karmalkar and K. N. Bhat, "The shifted-rectangle approximation for simplifying the analysis of ion-implanted MOSFETs and MESFETs," *Solid-State Electronics*, vol. 34, pp. 681-692, 1991.
48. L. Lukasiak and A. Jakubowski, "The influence of nonuniform doping profile on I - V characteristics of MOS transistors," *IEEE Transactions on Electron Devices*, vol. 40, pp. 453-455, February 1993.
49. K. Lee, M. Shur, T. A. Fjeldly, and T. Ytterdal, *Semiconductor Device Modeling for VLSI*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
50. U. Cilingiroglu, *Systematic Analysis of Bipolar and MOS Transistors*, Artech House, Boston, 1993.
51. N. Arora, *MOSFET Models for VLSI Circuit Simulation—Theory and Practice*, Springer-Verlag, Vienna, 1993.
52. T. Skotnicki, C. Denat, P. Senn, G. Merckel, and B. Hennion, "A new analog/digital CAD model for sub-halfmicron MOSFETs," *Proceedings IEEE 1994 International Electron Devices Meeting*, pp. 165-168, San Francisco, 1994.
53. J.-J. Maa and C.-Y. Wu, "A new simplified threshold-voltage model for n-MOSFET's with nonuniformly doped substrate and its application to MOSFET's miniaturization," *IEEE Transactions on Electron Devices*, vol. 42, p. 1487, August 1995.
54. C. Lallement, M. Bucher, and C. Enz, "Modeling and characterization of non-uniform substrate doping," *Solid-State Electronics*, vol. 41, pp. 1857-1861, 1997.
55. F. Gámiz, J. A. L.-Villanueva, J. B. Roldán, and J. E. Carceller, "Influence of the doping profile on electron mobility in a MOSFET," *IEEE Transactions on Electron Devices*, vol. 43, pp. 2023-2025, November 1996.
56. J. R. Brews, "Sensitivity of subthreshold current to profile variations in long-channel MOSFET's," *IEEE Transactions on Electron Devices*, vol. 43, p. 2614, December 1996.
57. C. McAndrew, private communication.
58. N. Rao, private communication.
59. J. R. Edwards and G. Mar, "Depletion-mode IGFET made by deep ion implantation," *IEEE Transactions on Electron Devices*, vol. ED-20, pp. 283-289, 1973.
60. J. S. T. Huang, "Characteristics of a depletion-mode IGFET," *IEEE Transactions on Electron Devices*, vol. ED-20, pp. 513-515, 1973.
61. T. W. Sigmon and R. Swanson, "MOS threshold shifting by ion implantation," *Solid-State Electronics*, vol. 16, pp. 1217-1232, 1973.

62. A. M. Mohsen and F. J. Morris, "Measurements on depletion-mode field-effect transistors and buried channel MOS capacitors for the characterization of bulk transfer charge-coupled devices," *Solid-State Electronics*, vol. 18, pp. 407-416, 1975.
63. J. R. Verjans and R. J. Van Overstraeten, "Electrical characteristics of boron-implanted n-channel MOS transistors for use in logic circuits," *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 862-868, 1975.
64. J. S. T. Huang and G. W. Taylor, "Modeling of an ion-implanted silicon-gate depletion mode IGFET," *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 995-1001, 1975. See also W. Marciniak and H. Madura, "Comments on the Huang and Taylor model of ion-implanted silicon-gate depletion-mode IGFET," *Solid-State Electronics*, vol. 28, pp. 313-315, 1985.
65. P. E. Schmidt and M. B. Das, "D.C. and high-frequency characteristics of built-in channel MOS-FETs," *Solid-State Electronics*, vol. 21, pp. 495-505, 1978.
66. G. R. Mohan Rao, "An accurate model for a depletion mode IGFET used as a load device," *Solid-State Electronics*, vol. 21, pp. 711-714, 1978.
67. R. A. Haken, "Analysis of the deep depletion MOSFET and the use of the dc characteristics for determining bulk-channel charge coupled device parameters," *Solid-State Electronics*, vol. 21, pp. 753-761, 1978.
68. T. E. Hendrikson, "A simplified model for subpinchoff condition in depletion mode IGFET's," *IEEE Transactions on Electron Devices*, vol. ED-25, pp. 435-441, 1978.
69. Y. A. El-Mansy, "A nonlinear CAD model for the depletion-mode IGFET," *Technical Digest, International Electron Devices Meeting*, Washington, D.C., pp. 20-25, 1978.
70. K. Nishiuchi, H. Oka, T. Nakamura, H. Ishikawa, and M. Shinoda, "A normally-off type buried channel MOSFET for VLSI circuits," *Technical Digest, International Electron Devices Meeting*, Washington, D.C., p. 26, 1978.
71. F. H. Gaensslen and K. C. Jaeger, "Temperature dependent threshold behavior of depletion mode MOSFETs," *Solid-State Electronics*, vol. 22, pp. 423-430, 1979.
72. M. Wordeman "Characterization of depletion-mode MOSFETS," *Digest International Electron Devices Meeting*, Washington, D.C., 1979, pp. 26-29.
73. G. Baccarani, F. Landini, and B. Ricco, "Depletion-mode MOSFET model including a field-dependent surface mobility," *IEE Proceedings*, vol. 127, part I, pp. 62-66, 1980.
74. Y. A. El-Mansy, "Analysis and characterization of the depletion-mode IGFET," *IEEE Journal of Solid-State Circuits*, vol. SC-15, pp. 331-340, 1980.
75. N. Ballay and B. Baylac, "Analytical modelling of depletion-mode MOSFET with short- and narrow-channel effects," *IEE Proceedings*, vol. 127, part I, pp. 225-230, December 1981.
76. M. R. Wordeman and R. H. Dennard, "Threshold voltage characteristics of depletion-mode MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-28, pp. 1025-1030, 1981.
77. U. Ohno and Y. Okuto, "Electron mobility in n-channel depletion type MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 190-194, 1982.
78. R. Ratnam and A. B. Bhattacharyya, "Accumulation-punchthrough model of operation of buried-channel MOSFET's," *IEEE Electron Device Letters*, EDL-3, pp. 203-204, 1982.
79. T. Yamaguchi and S. Morimoto, "Analytical model and characteristics of small geometry buried-channel depletion MOSFETs," *IEEE Journal of Solid-State Circuits*, vol. SC-18, pp. 784-793, 1983.
80. S. Haque-Ahmed and C. A. T. Salama, "Depletion mode MOSFET modelling for CAD," *IEE Proceedings*, vol. 130, part I, pp. 281-286, 1983.
81. J. S. T. Huang, J. W. Schrankler, and J. S. Kueng, "Short-channel threshold model for buried-channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 1889-1895, December 1984.
82. F. M. Klaassen and W. Hes, "Compensated MOSFET devices," *Solid-State Electronics*, vol. 28, pp. 359-373, 1985.
83. D. Ma, "A physical and SPICE-compatible model for the MOS depletion device," *IEEE Transactions on Computer-Aided Design*, vol. CAD-4, pp. 349-356, July 1985.
84. M.-W. Chiang, J. Choma, Jr., and C. Kao, "A simulation method to completely model the various transistor I-V operational modes of long channel depletion MOSFET's," *IEEE Transactions on Computer-Aided Design*, vol. CAD-4, pp. 322-328, July 1985.
85. G. J. Hu and R. H. Bruce, "Design tradeoffs between surface and buried channel FETs," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 584-588, March 1985.

86. C. Turchetti and G. Masetti, "Analysis of the depletion-mode MOSFET including diffusion and drift currents," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 773–782, April 1985.
87. C.-Y. Wu and K. C. Hsu, "Mobility models for the I-V characteristics of buried-channel MOSFETs," *Solid-State Electronics*, vol. 28, pp. 917–923, 1985.
88. K. C.-K. Weng, P. Yang, and J.-H. Chern, "A predictor/CAD model for buried-channel MOS transistors," *IEEE Transactions on Computer-Aided Design*, vol. CAD-6, pp. 4–16, January 1987.
89. S. Karmalkar and K. N. Bhat, "The correct equivalent box representation for the buried layer of BC MOSFET's in terms of the implantation parameters," *IEEE Electron Device Letters*, vol. EDL-8, pp. 457–459, October 1987.
90. Z. Yu and X. Zhao, "A semi-analytical approach to the evaluation of threshold voltage in depletion MOS's with nonuniformly doped substrates," *IEEE Transactions on Electron Devices*, vol. 35, p. 993, July 1988.
91. S. W. Tarasewicz and C. A. T. Salama, "Threshold voltage characteristics of ion-implanted depletion MOSFETs," *Solid-State Electronics*, vol. 31, pp. 1441–1446, 1988.
92. C. D. Parikh and J. Vasi, "Modeling of a depletion-mode MOSFET," *Solid-State Electronics*, vol. 30, pp. 699–703, 1987. See also comments by C. Turchetti and G. Masetti, *ibid.*, vol. 31, pp. 1747–1748, 1988, and authors' response, *ibid.*, p. 1749.
93. M. J. Van de Tol and S. G. Chamberlain, "Buried-channel MOSFET model for SPICE," *IEEE Transactions on Computer-Aided Design*, vol. CAD-10, pp. 1015–1035, 1991.
94. Y. Yin and J. A. Cooper, Jr., "Simple equations for the electrostatic potential in buried-channel MOS devices," *IEEE Transactions on Electron Devices*, vol. 39, pp. 1770–1772, July 1992.
95. B.-j. Moon, C.-k. Park, K.-m. Rho, K. Lee, M. Shur, and T. A. Fjeldly, "Analytical model for *p*-channel MOSFET's," *IEEE Transactions on Electron Devices*, vol. 38, pp. 2632–2646, December 1991.
96. C. Bulucea and D. Kerr, "Threshold voltage control in buried-channel MOSFETs," *Solid-State Electronics*, vol. 41, pp. 1345–1354, 1997.

PROBLEMS

- 5.1. Develop an algorithm which can be used by a computer to determine the drain current of a *p*-substrate-*p*-implant long-channel device given the model parameters and V_{GB} , V_{DB} , and V_{SB} . Assume the device is in strong inversion.
- 5.2. (Warning: this is a long problem.) This problem considers the development of the simplified quadratic model discussed in Sec. 5.2.4. The ideas involved are similar to the ones in Sec. 4.5.3.
 - (a) Show that the expression for Q'_B , (5.2.9) and (5.2.15), can be approximated by using

$$-\frac{Q'_B}{C'_{ox}} = \gamma_1 \sqrt{\phi_{01} + V_{SB}} + (\alpha_{I1} - 1)(V_{CB} - V_{SB}) \quad V_{SB} < V_{CB} \leq V_I$$

$$-\frac{Q'_B}{C'_{ox}} = \frac{qM}{C'_{ox}} + \gamma_2 \sqrt{\phi_{02} + V_{SB}} + (\alpha_{I2} - 1)(V_{CB} - V_{SB}) \quad V_I \leq V_{SB} < V_{CB}$$

$$-\frac{Q'_B}{C'_{ox}} = \frac{qM}{C'_{ox}} + \gamma_2 \sqrt{\phi_{02} + V_I} + (\alpha_{I2} - 1)(V_{CB} - V_I) \quad V_{SB} < V_I < V_{CB}$$

where α_{I2} is, until further notice, an empirical parameter [see part (g)], and α_{I1} is given by

$$\alpha_{I1} = 1 + \gamma_1 \frac{\sqrt{\phi_{01} + V_I} - \sqrt{\phi_{01} + V_{SB}}}{V_I - V_{SB}}$$

Plot the exact and approximate $-Q'_B/C'_{ox}$ (in analogy to Fig. 4.16) for all cases, and show that the first and third expressions for $-Q'_B/C'_{ox}$ give the same value at the boundary $V_{CB} = V_f$, and this value is exact. Comment as to why these are useful properties of the approximate expressions.

- (b) Propose empirical values for α_{I2} (in analogy to those for α in Sec. 4.5.3) and show that $\alpha_{I2} < \alpha_{I1}$. Show that for best accuracy of the equations given above, α_{I2} should be made dependent on V_{SB} for $V_f \leq V_{SB} < V_{CB}$ and on V_f for $V_{SB} < V_f < V_{CB}$.
 - (c) Discuss the numerical problem that the above expression for α_{I1} has at $V_{SB} = V_f$, and show that it can be rewritten⁵⁷ as in (5.2.45a), in which case no numerical problem occurs.
 - (d) Use the above approximations to derive the model in (5.2.39) with (5.2.44).
 - (e) Derive the expressions for V_p given in (5.2.47) to (5.2.49).
 - (f) Check the model for continuity of the current and its derivatives with respect to all bias voltages at appropriate critical points. Explain the choice of α_2 as in (5.2.45b), for reasons of continuity of dI_{DS}/dV_{SB} at $V_{SB} = V_f$, and comment on the ensuing effect on the accuracy of I_{DS} .
 - (g) Show that, if a value different from the one suggested above is used for α_{I1} , dI_{DSN}/dV_{DB} can be discontinuous at $V_{DB} = V_f$.
- 5.3. Show that, with $V_T(V_{SB})$ as given by (5.2.30), both $V_T(V_{SB})$ and dV_T/dV_{SB} are continuous at $V_{SB} = V_f$, and⁵⁸

$$\left. \frac{dV_T}{dV_{SB}} \right|_{V_{SB}=V_f} = \frac{\epsilon_s}{\epsilon_{ox}} \frac{t_{ox}}{d_f} \approx 3 \frac{t_{ox}}{d_f}$$

- 5.4. The channel of an NMOS transistor with $t_{ox} = 100 \text{ \AA}$ and a p substrate with $N_A = 3 \times 10^{15} \text{ cm}^{-3}$ is implanted with *acceptors* using an effective dose of 10^{12} cm^{-2} . The value of d_f is $0.2 \text{ }\mu\text{m}$. Assume $W/L = 10$, $\mu = 700 \text{ cm}^2/(\text{V} \cdot \text{s})$, $V_{FB1} = -0.8 \text{ V}$ and $\phi_{01} = 0.7 \text{ V}$. Plot (a) $V_T(V_{SB})$ vs. V_{SB} ; (b) I_{DS} vs. V_{DS} with V_{GS} as a parameter, for $V_{SB} = 0 \text{ V}$; and (c) $\sqrt{I_{DS}}$ vs. V_{GS} in saturation for $V_{SB} = 0 \text{ V}$ and $V_{SB} = 2 \text{ V}$. Choose appropriate voltage ranges so that all pertinent effects are evident on the plots.
- 5.5. Find an accurate expression for $V_T(V_{CB})$ in (5.3.9). Show that for $N_I \gg N_{AB}$ it reduces to (5.3.10).
- 5.6. The channel of an NMOS transistor with $t_{ox} = 100 \text{ \AA}$ and a p substrate with $N_A = 5 \times 10^{15} \text{ cm}^{-3}$ is implanted with *donors* using an effective dose of 10^{12} cm^{-2} ; the value of d_f is $0.1 \text{ }\mu\text{m}$. Assume $W/L = 5$, $\mu_s = 600 \text{ cm}^2/(\text{V} \cdot \text{s})$, $\mu_B = 800 \text{ cm}^2/(\text{V} \cdot \text{s})$, and $V_{FB} = -0.85 \text{ V}$. Using the model of Table 5.1, plot (a) I_{DS} vs. V_{DS} with V_{GS} as a parameter, for $V_{SB} = 0 \text{ V}$; (b) I_{DS} vs. V_{GS} for $V_{DS} = 0.1 \text{ V}$, for $V_{SB} = 0 \text{ V}$ and $V_{SB} = 2 \text{ V}$. Choose appropriate bias ranges so that all pertinent effects are evident on the plots.
- 5.7. For the device of Prob. 5.6, what is the maximum value of d_f for which turnoff at $V_{SB} = 0$ can be guaranteed with a sufficiently negative V_{GS} (i.e., so that the problem illustrated in Figs. 5.13 and 5.17 does not occur)?
- 5.8. Find the values for V'_{DS1} and V'_{DS2} used in Fig. 5.14 by using equations from the main text. Compare the resulting expressions to the corresponding approximate relations in Table 5.1 and comment.

- 5.9.** Consider an n -implant- p -substrate device in surface depletion operation. It has been proposed⁶⁴ to approximate Q'_T (see Fig. 5.12*b*) by

$$Q'_T \approx \bar{C}'(V_{GB} - V_{FB} - \phi_{bi} - V_{CB})$$

where \bar{C}' is an approximate parameter having the dimensions of capacitance. Show that the above relation can be derived from (5.3.7) by performing an appropriate Taylor series expansion of the square root, keeping the first two terms in it, and adjusting the coefficient of the first-order term for better accuracy. Comment on the limits of validity of such an approximation.

- 5.10.** Using the expression in Prob. 5.9, show that the drain current for surface depletion operation can be approximated by an equation in the form of (4.5.2) if new "effective" values are defined for V_{FB} and ϕ_0 .⁶⁴ Compare qualitatively the resulting expression to the model in Table 5.1.
- 5.11.** Prove (5.2.15) using basic electrostatics (Appendix B). Plot the charge density, the electric field, and the potential as a function of y . Assume that the potential is 0 deep in the substrate, and that the total potential drop across the entire depletion region is $\phi_{01} + V_{CB}$. Find the depth of the depletion region in the substrate, and from that find Q'_{B2} .

CHAPTER 6

SMALL- DIMENSION EFFECTS

This chapter has been revised and expanded by D. A. ANTONIADIS,
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.1 INTRODUCTION

In the previous chapter we assumed that the transistors under consideration had a channel sufficiently long and wide, so that “edge” effects along the four sides of the channel could be neglected. This allowed us to assume that the electric field lines were everywhere perpendicular to the surface (i.e., they had components only along the y direction), and we performed what is called a *one-dimensional* analysis by using the gradual-channel approximation. The equations we derived based on such assumptions fail to characterize adequately devices with short and/or narrow channels. If the channel is short (i.e., L is not much larger than the sum of the source and drain depletion widths), a significant part of the electric field lines will have components along both the y and the x directions, the latter being the direction along the channel’s length. Thus, a *two-dimensional* analysis will be needed. If the channel is instead narrow (i.e., W is not much larger than the depletion region depth under the gate), a significant part of the field lines will have components along the y and z directions, the latter being the direction along the channel’s width. Again, a two-dimensional analysis must be employed. If the channel is short *and* narrow, field lines will, in general, have components along the x , y , and z directions; now a *three-dimensional* analysis becomes necessary.

In addition to its increased dimensionality, the *magnitude* of the electric field in very small geometry MOSFETs tends to increase because typically terminal voltages

are not scaled-down in proportion to device geometrical dimensions. The increase of the longitudinal electric field in the channel of MOSFETs does not translate entirely to increased current because, as we will see in this chapter, the carrier velocity in the channel saturates to a finite value, which limits the current. However, the random energy of carriers, which in analogy to gas molecules can be characterized by a carrier "temperature," increases in the presence of high longitudinal fields leading to what are called "hot-carrier" phenomena that affect the device operation and also produce operational "wearout" that limits the device lifetime. These phenomena, along with others associated with the increase of the vertical electric field that lead to additional device performance degradation, are described in this chapter.

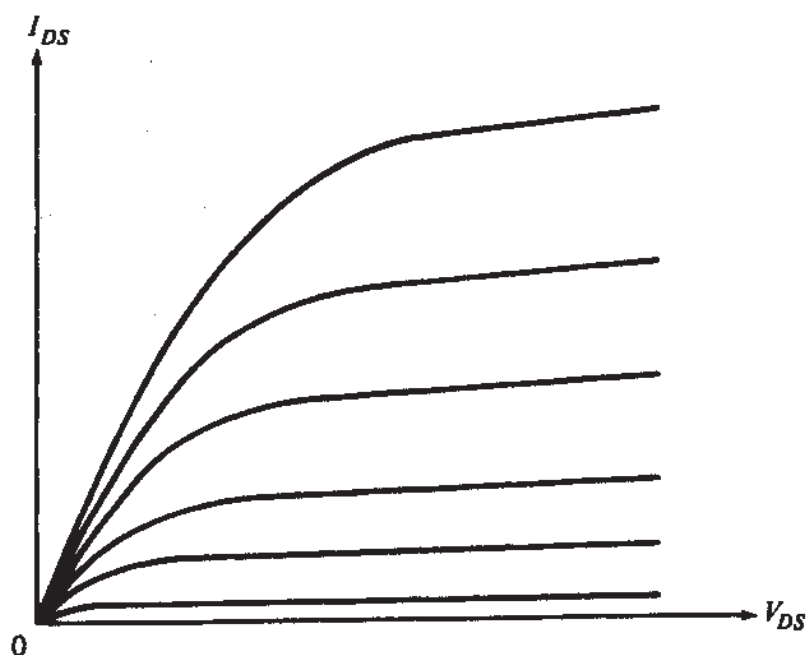
Two- and three-dimensional analyses can be carried out numerically with the help of a computer.¹⁻¹⁵ However, such analyses, albeit accurate, do not provide a simple model for efficient calculation. Thus, empirical approximations and semiempirical approaches have been used to provide useful simplification. In these, usually the complex two- or three-dimensional phenomena are broken down into simple, separate phenomena examined one at a time. A number of simplifying assumptions are then made, *which are sometimes difficult to justify rigorously*, and relatively simple relations are derived. Often such techniques are characterized by an attempt to maintain the general form of the I - V relations for the long- and wide-channel devices, and to "stretch" these relations by modifying them somewhat so that they can be used in the case of short and/or narrow channels. What is considered to justify these empirical approaches is their success in simulating the behavior observed experimentally. Although some of these approaches necessarily lack rigor and elegance, they have often been helpful where more precise work failed to give computationally efficient models. Some representative examples of the semiempirical modeling process, which has largely been limited to the strong-inversion region, will be presented in this chapter.

6.2 CHANNEL LENGTH MODULATION

As mentioned in Sec. 4.5, in the saturation region the I_{DS} - V_{DS} characteristics are not exactly parallel to the horizontal axis but have, instead, a positive slope.¹⁶⁻⁴⁰ Other things being equal, this slope has been found to be larger for shorter channels, and may be easily noticeable, as shown in Fig. 6.1. Historically, this phenomenon was the first "short-channel effect" to be studied. It was not originally classified as such, partly because its investigation began well before various other short-channel effects were recognized and named that way, and partly because it can play an important role in circuit work even in devices with long channels (e.g., 10 μm or longer).†

Two-dimensional analyses of the region near the drain in saturation present a very complicated picture.^{20,31} Field lines emanate from the drain n^+ region and terminate on

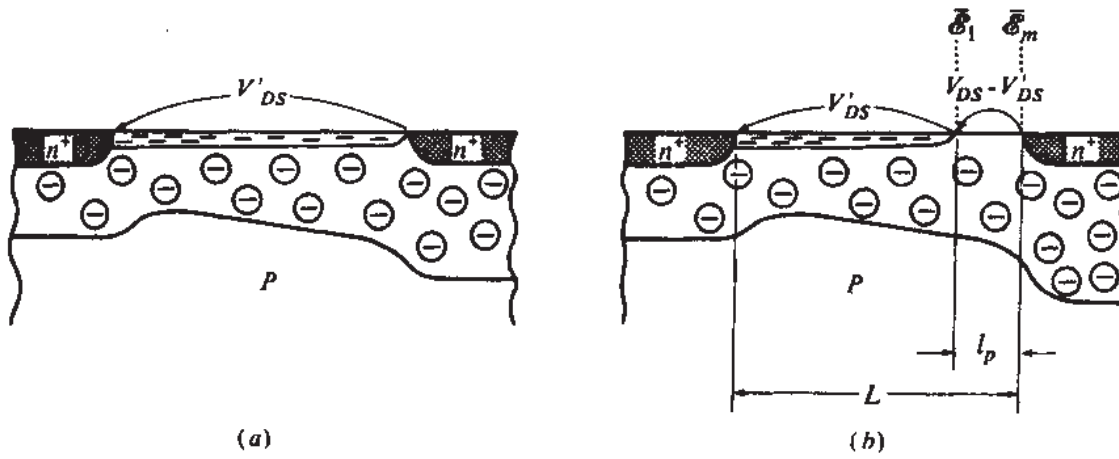
†An example is the analog "CMOS inverter." The small-signal gain of such a circuit turns out to be inversely proportional to the sum of the I_{DS} - V_{DS} slopes of two devices in saturation. The models of Sec. 4.5 would thus predict infinite gain!

**FIGURE 6.1**

Transistor characteristics in the presence of channel length modulation.

points in the channel. Of these some are nearly horizontal, but others can start from the bottom of the n^+ region with a downward direction, curve gradually upward, and terminate on the inversion layer. In addition, the field lines from the gate to the channel and from the channel to the bulk curve in a complicated manner near the drain. The concentration of electrons decreases toward the drain, and these electrons are pushed away from the surface and into the bulk. Thus, one can think of the channel as curving downward in the vicinity of the drain and of the current flowing in a "subsurface" path there. From such a picture one can suspect that the drain junction depth can affect the current value and, in fact, this is the case.

The above complicated picture has so far failed to provide simple analytical models. However, many semiempirical formulations have been derived.^{16-19,21-48} Sometimes, to obtain approximate but manageable analytical results for the behavior in saturation, a greatly simplified picture is used; this will now be described. Figure 6.2a shows the semiconductor part of a transistor with $V_{DS} = V'_{DS}$, with V'_{DS} being the value at which "pinchoff" is assumed to occur at the drain end of the channel, as described in Sec. 4.5.3. As explained there, under this assumption $|Q'_I|$ at the drain end has a small nonzero value, which is much smaller than the magnitude of the depletion region charge per unit area at that end. If now V_{DS} is increased above V'_{DS} , $|Q'_I|$ at the drain end will decrease below the above value. The pinched-off tip of the inversion layer will now move to the left, as shown in Fig. 6.2b, and the region between it and the drain n^+ region is approximated by a depletion region. In such a picture, the assumption of a depletion region is only approximate and does *not* imply zero current. The reader might be familiar with pn junctions and bipolar transistors in which large currents can flow through depletion regions. However, since $|Q'_I|$ is small in this region, one must assume that the electrons travel at high speeds in it in order for a considerable value of I_{DS} to be possible.

**FIGURE 6.2**

Channel (a) at pinchoff; (b) above pinchoff. \mathcal{E}_1 is the magnitude of the horizontal electric field at the end of the inversion layer (i.e., the beginning of the pinchoff region) and \mathcal{E}_m is the maximum electric field magnitude which develops near the n^+ junction.

In Fig. 6.2b the channel cannot support more voltage than V'_{DS} since it becomes pinched off when the voltage across it reaches that value. The *excess* voltage $V_{DS} - V'_{DS}$ must then be dropped between the drain and the tip of the channel. Such a nonzero voltage can only exist over a region of nonzero length l_p , as shown. If V_{DS} is raised still further, more excess voltage must be dropped across the depletion region. To support this voltage, the region must widen, and the inversion layer will shrink somewhat in length. This is referred to as *channel length modulation* (CLM). If the electron concentration in the depletion region is very small, we can assume that practically all the charge in the region is due to ionized acceptor atoms, with volume concentration N_A . Let us denote by \mathcal{E}_1 the magnitude of the channel field intensity in the x direction at the left end of the depletion region. Then, assuming that the field in the depletion region near the surface is approximately horizontal, it can be shown by using Poisson's equation (1.2.13) that the length l_p of the depletion region can be related to the potential $V_{DS} - V'_{DS}$ across it by (Prob. 6.1)

$$l_p = \sqrt{\frac{2\epsilon_s}{qN_A}} \left[\sqrt{\phi_D + (V_{DS} - V'_{DS})} - \sqrt{\phi_D} \right] \quad (6.2.1)$$

where

$$\phi_D = \frac{\epsilon_s \mathcal{E}_1^2}{2qN_A} \quad (6.2.2)$$

According to this simple model, the current conduction mechanism is as follows. Electrons enter the source end and travel along the channel; at the pinchoff point, they find themselves in what is practically a depletion region. As seen in Fig. 6.2b, the direction of the field in that region is such as to sweep these electrons through the depletion region to the drain. Note that the above one-dimensional model

totally ignores the influence of the gate on the electric field near the drain, which can be important.^{20,22,29,38}

In a different formulation,^{23,24} pinchoff is taken to occur when the field near the drain becomes high enough to cause velocity saturation† (assuming a simplified model for electron velocity in which velocity saturation is attained abruptly at a finite value of field intensity). For V_{DS} above the value corresponding to this point, the electrons are assumed to travel in the depletion region at maximum velocity. Such a formulation results again in (6.2.1) and (6.2.2), with \mathcal{E}_1 taken to be a field value above which the electron velocity is assumed saturated (e.g., for electrons this value is 8×10^3 to 3×10^4 V/cm; see Sec. 6.5 for more details). In some treatments, pinchoff is taken to correspond to the point at which the inversion layer “dives” below the interface, as predicted by two-dimensional simulations. This point can be defined more precisely as the point at which the surface electric field component in the vertical direction becomes zero.³¹ For the orientation of the device in Fig. 6.2*b*, the gate field “pulls” the inversion layer toward the interface to the left of the pinchoff point and “pushes” it below the interface to the right of this point. Such a formulation can again result in (6.2.1) and (6.2.2), but with an \mathcal{E}_1 value which must be found from a numerical solution. In practice, independently of which definition is used for pinchoff, one usually ends up adjusting the value of \mathcal{E}_1 in (6.2.2) for a best fit to experimental results. A single value of \mathcal{E}_1 is sometimes used for simplicity, which lies between 10^4 and 2×10^5 V/cm.

Let us now consider the effect of these phenomena on the drain current. At $V_{DS} = V'_{DS}$, the current I'_{DS} is usually computed from the nonsaturation equations, which are assumed to be valid up to this point. From such computations [see, for example, (4.5.12) with (4.5.2)] we see that

$$I'_{DS} = (\text{const}) \frac{1}{L} \quad (6.2.3)$$

where the constant of proportionality (const) depends on V_{GS} and V_{SB} .

Consider now some value of V_{DS} larger than V'_{DS} , and let the corresponding current be I_{DS} . This current can be calculated by considering part of the channel that is *not* pinched off in Fig. 6.2*b*. For that part, the situation is the same as in Fig. 6.2*a*, save for the fact that the role of L is now played by $L - l_p$; thus, similar to (6.2.3), we will have

$$I_{DS} = (\text{const}) \frac{1}{L - l_p} \quad (6.2.4)$$

† Velocity saturation was introduced in Sec. 1.3. Its effect on MOS transistor characteristics will be considered in detail in Sec. 6.5.

From the above two equations we obtain:

$$I_{DS} = I'_{DS} \frac{L}{L - l_p} \quad (6.2.5a)$$

or

$$I_{DS} = \frac{I'_{DS}}{1 - l_p/L} \quad (6.2.5b)$$

If $l_p/L \ll 1$, this can be approximated by

$$I_{DS} \approx I'_{DS} \left(1 + \frac{l_p}{L} \right) \quad (6.2.5c)$$

This equation is often preferred to (6.2.5b) in computer simulation models [note that, for some combinations of device parameters and terminal voltages, the denominator of (6.2.5b) could become zero, although this is unlikely].

The value of l_p predicted by (6.2.1) is

$$l_p = \frac{B_1}{\sqrt{N_A}} \left[\sqrt{\phi_D + (V_{DS} - V'_{DS})} - \sqrt{\phi_D} \right] \quad (6.2.6)$$

where $B_1 = (2\epsilon_s/q)^{1/2}$. In practice, however, the values of this constant and of ϕ_D can be chosen empirically for best matching between the resulting expressions for I_{DS} and experimental data. Such empirical adjustments will be considered in several places in this chapter. They become necessary in oversimplified descriptions of complicated phenomena, in order to help the accuracy of the resulting oversimplified results. Although empirical adjustments are mentioned in this book in order to familiarize the reader with common practice, it should be stressed that they should be used only as a last resort, after all attempts to find a correct physical description with reasonable complexity, have failed. See also Chap. 10.

An even simpler formulation¹⁹ results in $\phi_D = 0$. However, such a formulation is based on the inaccurate assumption that the field E_1 at the pinched off point is zero. It has the additional problem that, as V_{DS} is reduced toward V'_{DS} , it predicts a slope dI_{DS}/dV_{DS} which tends to infinity.

The error in the saturation current predicted by (6.2.5) and (6.2.6) is acceptable for some applications. However, it is possible for the error in I_{DS} to be small, while at the same time the error in the slope dI_{DS}/dV_{DS} is large, especially if dI_{DS}/dV_{DS} is small in the first place. This is discussed in detail in Sec. 8.2.2. In cases where accurate prediction of dI_{DS}/dV_{DS} is essential (e.g., in analog circuit design), the above model is not adequate. One may have to go to more precise (and much more complicated) analyses, which include the effect of the gate on the field in the pinchoff region as well as the nonzero charge and shape of the inversion layer in that region.

Such analyses can be expected to lead to expressions for l_p of the following general form

$$l_p = f_l(L, N_A, t_{ox}, d_j, \phi_{MS}, Q_o, V_{DS}, V_{GS}, V_{SB}) \quad (6.2.7)$$

where d_j is the drain junction depth. The derivation of results in the above form requires two-dimensional or pseudo-two-dimensional analyses and, again, the use of certain simplifications (recall that the whole idea behind l_p is a simplification). Several such analyses have been presented in the literature,^{20-22,27-31,34,40,42,43} although not all of them take every effect mentioned into account.

According to one of the most accepted pseudo-two-dimensional analyses,^{31,43} the pinchoff region length, can be written as

$$l_p = l_a \ln \frac{[(V_{DS} - V'_{DS})/l_a] + \mathcal{E}_m}{\mathcal{E}_1} \quad (6.2.8)$$

where \mathcal{E}_m is the maximum magnitude of the x -directed electric field in the pinchoff region (see Fig. 6.2b) given by^{31,43}

$$\mathcal{E}_m = \sqrt{\frac{(V_{DS} - V'_{DS})^2}{l_a^2} + \mathcal{E}_1^2} \quad (6.2.9)$$

with \mathcal{E}_1 the magnitude of the field in the beginning of the pinchoff region, and l_a is a characteristic length given by

$$l_a = \sqrt{\frac{\epsilon_s}{\epsilon_{ox}} t_{ox} d_j} \approx \sqrt{3 t_{ox} d_j} \quad (6.2.10)$$

In this model, \mathcal{E}_1 is taken to be equal to the field for electron or hole velocity saturation (denoted by \mathcal{E}_c in Sec. 6.5). The formulation in (6.2.8) to (6.2.10) has been shown to compare favorably with two-dimensional numerical device simulation, although a better (empirical) approximation for l_a appears to be⁴³ $l_a = (0.22 \text{ cm}^{1/6}) d_j^{1/2} t_{ox}^{1/3}$.

Instead of being represented by (6.2.9), \mathcal{E}_m is sometimes approximated by a function of the form^{88,148} $\mathcal{E}_1 + (\text{const}) [(V_{DS} - V'_{DS})/l_a]$. Using this in (6.2.8), we easily see that l_p is of the form

$$l_p = l_a \ln \left[1 + \frac{V_{DS} - V'_{DS}}{V_E} \right] \quad (6.2.11)$$

The parameter V_E , although it can in principle be expressed in terms of quantities given so far, is best determined experimentally. V_E is typically less than 1 V.

For *digital* circuit design, where dI_{DS}/dV_{DS} in saturation is often not important *per se*, the results mentioned in the preceding paragraphs are sometimes avoided for

reasons of computational speed. In fact, sometimes models simpler than (6.2.5) and (6.2.6) are used, predicting a first-degree dependence of I_{DS} on V_{DS} .²¹ To relate those models to the ones above, let us substitute (6.2.6) into (6.2.5c), expand the resulting expression in a Taylor series around $V_{DS} = V'_{DS}$, and keep only the first two terms. The result is (Prob. 6.2)

$$I_{DS} = I'_{DS} \left(1 + \frac{V_{DS} - V'_{DS}}{V_A} \right) \quad (6.2.12)$$

with

$$V_A = B_2 L \sqrt{N_A} \quad (6.2.13)$$

where B_2 is a constant of proportionality of the order of 10^{-3} to $2 \times 10^{-3} \text{ V} \cdot \text{cm}^{1/2}$. Equation (6.2.12) is plotted in Fig. 6.3a. The intercept with the horizontal axis is $-V_A + V'_{DS}$, and thus depends on V_{GS} through V'_{DS} . Another empirical model in use is³²

$$I_{DS} = I'_{DS} \left(1 + \frac{V_{DS} - V'_{DS}}{V_A + V'_{DS}} \right) \quad (6.2.14)$$

The plot of this equation intercepts the horizontal axis at $-V_A$, as indicated in Fig. 6.3b, independent of V'_{DS} . Thus, saturation curves for various V_{GS} values are all assumed to be straight lines which, if extended, cross one another at the same point on the horizontal axis. Such behavior might be approximately observed in real devices for some fabrication processes.

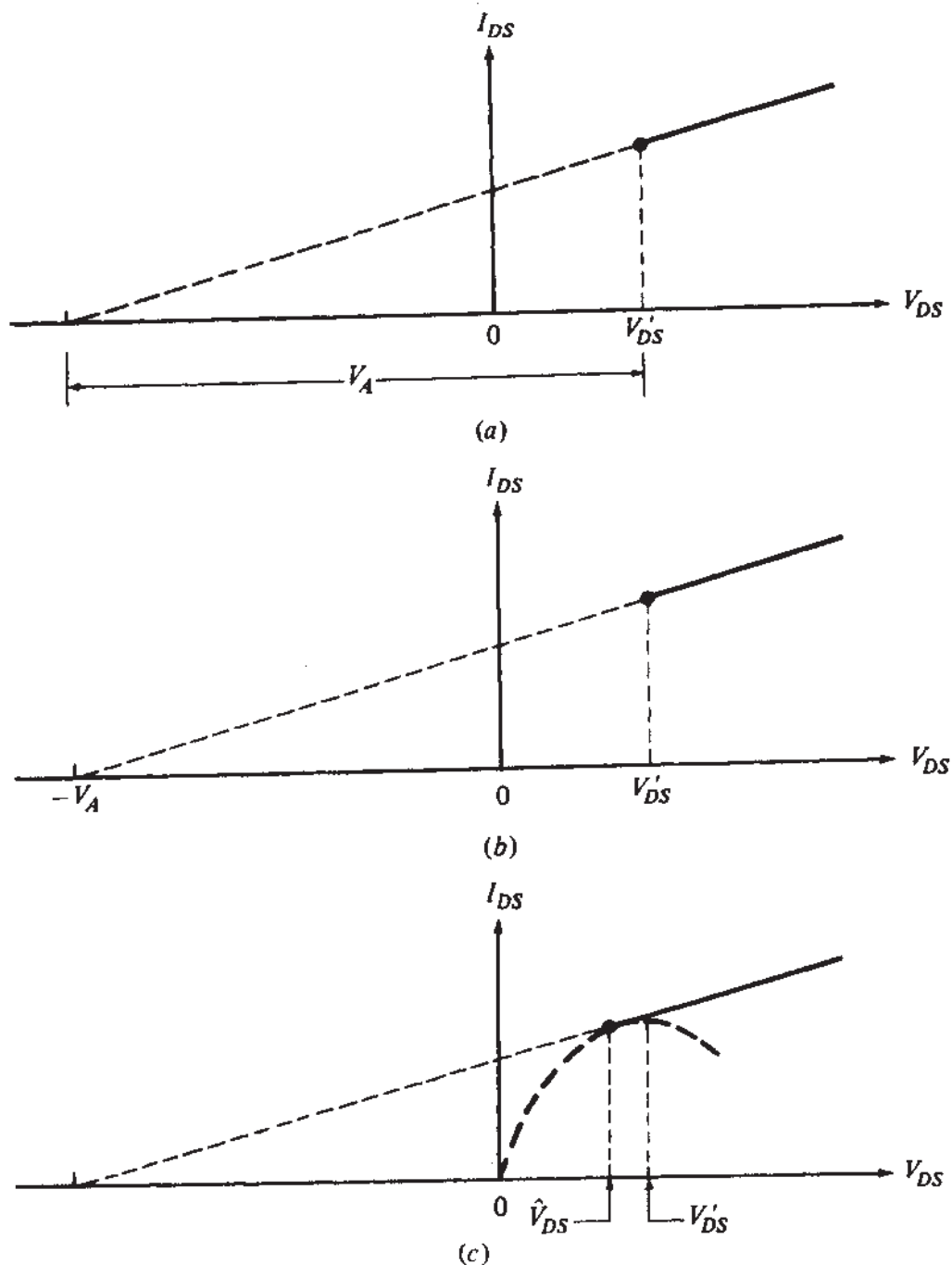
Note that the above saturation curves predict nonzero slope at V'_{DS} , whereas the nonsaturation equations predict zero slope at the same point. Such discontinuous slope behavior is unnatural and is also undesirable in the numerical algorithms used in the computer simulation of circuits. To avoid this problem, the saturation lines are drawn tangentially to the nonsaturation curves, as shown in Fig. 6.3c. Thus the limit between nonsaturation and saturation is defined slightly to the left of V'_{DS} ; this limit is shown by \hat{V}_{DS} in Fig. 6.3c.

Example 6.1. Let us develop a model based on the nonsaturation equation (4.5.37a) and the saturation equation (6.2.14), but using \hat{V}_{DS} instead of V'_{DS} due to the above considerations. We will have

$$I_{DS} = \frac{W}{L} \mu C'_{ox} \left[(V_{GS} - V_T) V_{DS} - \frac{\alpha}{2} V_{DS}^2 \right], \quad V_{DS} \leq \hat{V}_{DS} \quad (6.2.15)$$

and

$$I_{DS} = \hat{I}_{DS} \left(1 + \frac{V_{DS} - \hat{V}_{DS}}{V_A + \hat{V}_{DS}} \right), \quad V_{DS} > \hat{V}_{DS} \quad (6.2.16)$$

**FIGURE 6.3**

First-order modeling of the saturation region. (a) As provided by (6.2.12); (b) as provided by (6.2.14); (c) modification to ensure continuous slope at the transition point.

where \hat{I}_{DS} is the value of current at the upper limit of validity of (6.2.15):

$$\hat{I}_{DS} = \frac{W}{L} \mu C'_{ox} \left[(V_{GS} - V_T) \hat{V}_{DS} - \frac{\alpha}{2} \hat{V}_{DS}^2 \right] \quad (6.2.17)$$

The value of \hat{V}_{DS} required for continuous slope can now be obtained by equating dI_{DS}/dV_{DS} as obtained from (6.2.15) and (6.2.16). The result is (Prob. 6.3)

$$\hat{V}_{DS} = V_A \left[\sqrt{1 + \frac{2(V_{GS} - V_T)}{\alpha V_A}} - 1 \right] \quad (6.2.18)$$

From the graphical construction in Fig. 6.3c we expect that, as V_A approaches infinity, \hat{V}_{DS} should approach V'_{DS} . This is easily checked to be the case with (6.2.17) (Prob. 6.3): As V_A approaches infinity, \hat{V}_{DS} approaches the value of $(V_{GS} - V_T)/\alpha$, which is, indeed, the value of V'_{DS} for the model considered in the absence of channel length modulation [see (4.5.34)].

A method to combine the nonsaturation and saturation expressions into a single expression using appropriate "smoothing functions" has been described elsewhere.⁴⁰

We emphasize again that (6.2.12) and (6.2.14) are good only for approximate predictions of the drain current in saturation. In applications where accurate prediction of the *slope* dI_{DS}/dV_{DS} is required (e.g., analog circuits), these equations are totally inadequate.

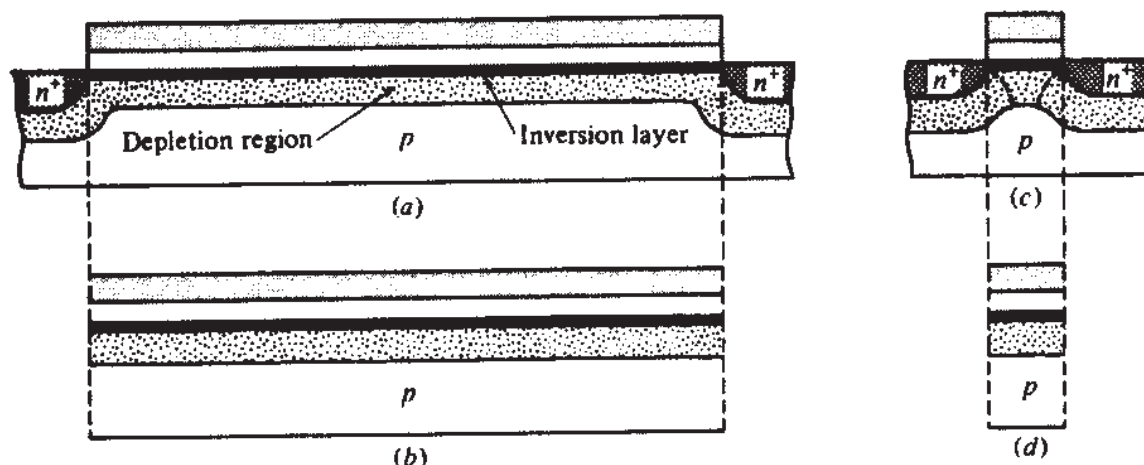
6.3 BARRIER LOWERING, TWO-DIMENSIONAL CHARGE SHARING, AND THRESHOLD VOLTAGE

6.3.1 Introduction

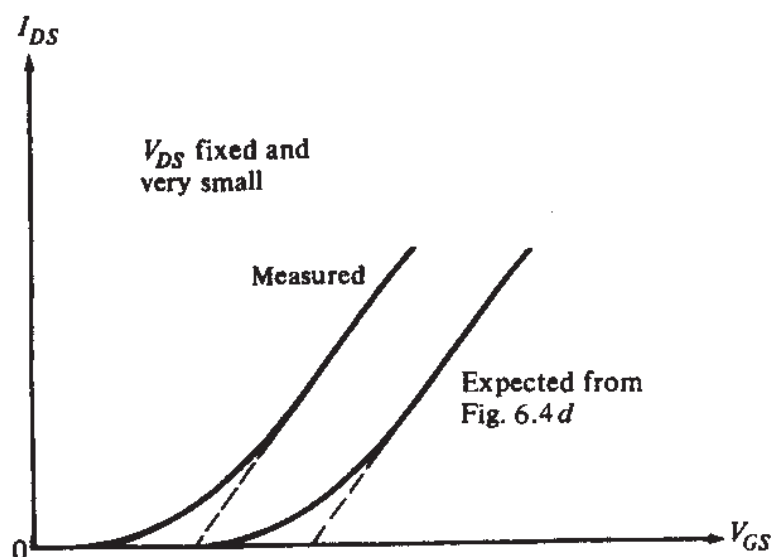
The short- and narrow-channel effects to be considered in this section are of such a nature that they can be approximately described by the equations developed in Sec. 4.5.3, provided that in them the threshold voltage is replaced by another quantity, called the *effective threshold voltage*. This quantity will be seen to depend on channel length, channel width, source-substrate voltage, and drain-source voltage.⁴⁹⁻¹⁰⁶ We will initially consider the cases of short channels and of narrow channels separately.

6.3.2 Short-Channel Devices

Let us review some of our assumptions for a *long*-channel device, shown in Fig. 6.4a; for simplicity, we assume here $V_{DS} = 0$. Neglecting the edge effects, as we have done so far, is equivalent to assuming that the situation between source and drain is the same as one would have with the source and drain removed, as shown in Fig. 6.4b, but with the channel somehow still communicating with the external world. Calculating Q'_I and then I_{DS} by using this assumption provides accurate results as long as the value of L is large. Consider now a short-channel device, as shown in Fig. 6.4c, made with the same process and biased with the same terminal voltages as the long-channel device. Here the edge effects practically extend over all of the channel. Neglecting these effects amounts to viewing the device as shown in Fig. 6.4d, which can hardly be expected to provide credible results. Indeed, assume a very small but nonzero V_{DS} , so that drain current can flow without appreciably disturbing the picture in the channel in Fig. 6.4c. It has been found experimentally that the V_{GS} value required to produce a given I_{DS} value is smaller than what would have been expected from the picture in Fig. 6.4d, as shown in Fig. 6.5. To get some intuitive feeling for the reasons behind this, one can use a number of viewpoints found in the literature. One of these uses the concept of *barrier lowering*,^{60,66} referring to the following. The effect of decreasing L is to deplete more of the region under the inversion layer; this is easily visualized if the source and drain are imagined to approach one another as shown in

**FIGURE 6.4**

(a) A long-channel transistor; (b) the channel of (a) with edge effects neglected; (c) a short-channel resistor; (d) the channel of (c) with edge effects neglected.

**FIGURE 6.5**

Drain current vs. gate-source voltage for very small, fixed drain-source voltage.

Fig. 6.4c. The deeper depletion region is accompanied by a larger surface potential,[†] which makes the channel more attractive for electrons. Thus the device can conduct more current than what would be predicted from long-channel theory for a given V_{GS} . Since, in long-channel theory, the drain current is an increasing function of $V_{GS} - V_T$, the long-channel equations may be made to describe artificially the current increase if V_T is replaced by a smaller quantity \hat{V}_T , which will be called the *effective threshold voltage*. This quantity decreases as L is decreased. For devices with the minimum possible channel length in a given technology, \hat{V}_T can be smaller than V_T by 50 to 200 mV. Notice that in the short-channel device of Fig. 6.4c the depletion

[†]In an energy band formulation, an increase in surface potential corresponds to a decrease in the potential energy for electrons. The potential energy "barrier" to the entrance of electrons in the channel is lowered; hence the name *barrier lowering* is used to describe these phenomena.

region under the channel can also be widened (and the corresponding surface potential can be increased) by raising the potential of the drain; this will further increase I_{DS} above the value expected from other considerations. Hence, the effective threshold \hat{V}_T need also be made a decreasing function of V_{DS} . Unfortunately, it is not easy to come up with simple analytical results by using the barrier-lowering concept.

An alternative description goes as follows. The charge in the channel region is influenced by field lines emanating from all nearby structures. Two such structures considered in the analysis of long-channel devices are, of course, the gate and the substrate ("back gate"). In short-channel devices the source and drain are so close to all points in the channel that they can affect the latter through their proximity just as the gate does. In effect, the source and drain now play something of a gate role in addition to their normal function. Field lines emanating from all four structures (gate, bulk, source, and drain) and terminating on charges in the channel must be considered for an accurate description of the device. Essentially, charge control in the channel is shared by all four structures. Bringing the source and drain regions closer to all points in the channel is similar to bringing the gate closer to the channel. Since source and drain help support the channel charge, the latter effect increases as L is decreased, for given V_{GS} and V_{SB} . The corresponding increase in drain current, however, is not predicted by long-channel theory and is modeled instead by using an effective threshold as explained above. Also, increasing the drain potential increases the inversion layer charge just as increasing the gate potential would. The above picture again leads us to expect an effective threshold which decreases with decreasing L and increasing V_{DS} .

This "charge-sharing" viewpoint has provided the basis for most of the analytical/empirical models for short-channel devices.^{49-53,57-59,61-65,67,70-74,78-80,82,83,85-88} We will expand on this idea below. But first we should warn the reader that charge-sharing models have been developed with the aim to obtain simple semiempirical expressions for describing very complex two-dimensional phenomena; in such developments, *it has not been possible to justify rigorously all steps*. To begin, let us assume that the equivalent interface charge Q'_o is zero for simplicity. In Figs. 6.4c and d, it can be imagined that one field line emanates from each positive charge q on the gate. Such a field line terminates either on an electron in the inversion layer or on an ionized acceptor in the depletion region, neglecting the "fringing" field lines terminating on the n^+ regions. In Fig. 6.4d, all of the depletion region charges are "imaged" on gate charges through connecting field lines. In the more realistic picture of Fig. 6.4c, however, this is not the case. Some of the field lines terminating on ionized acceptor atoms near the n^+ regions can be originating from ionized donor atoms in the n^+ regions (inside a thin depletion layer there). Thus, only part of the depletion region charge is imaged on the gate charge in this case. If we can assume that the gate charge in c and d is the same, and since some gate charges in c cannot be imaged on depletion charges, more of them are available to be imaged on inversion layer charges, which must, thus, increase to accept the extra field lines. The extra I_{DS} observed is attributed to this extra inversion layer charge, although, strictly speaking, one would have to know the spatial distribution of that charge before such a conclusion could be reached. A number of arbitrary assumptions are obviously involved in the above arguments (Prob. 6.6). Some of these assumptions are critically considered elsewhere.^{75,82}

Next, a large empirical step is taken: The short-channel device is assumed to behave as a fictitious device with a uniform depletion region, but with an effective

depletion region charge \hat{Q}_B , which has a smaller magnitude than the corresponding quantity Q_B in Fig. 6.4d. Because \hat{Q}_B and Q_B are defined for devices with uniform depletion regions and equal gate areas, we will have $\hat{Q}_B'/Q_B' = \hat{Q}_B/Q_B$, where the primes denote quantities per unit area. The above assumption for an effective charge \hat{Q}_B then implies that, instead of the long-channel threshold given by (3.4.14a),

$$V_T = V_{FB} + \phi_0 - \frac{Q_B'}{C'_{ox}} \quad (6.3.1a)$$

$$= V_{FB} + \phi_0 + \gamma \sqrt{\phi_0 + V_{SB}} \quad (6.3.1b)$$

we will have an effective threshold given by

$$\hat{V}_T = V_{FB} + \phi_0 - \frac{\hat{Q}_B'}{C'_{ox}} \quad (6.3.2a)$$

or

$$\boxed{\hat{V}_T = V_{FB} + \phi_0 + \frac{\hat{Q}_B}{Q_B} \gamma \sqrt{\phi_0 + V_{SB}}} \quad (6.3.2b)$$

We thus see that the above charge-sharing effect can be viewed as resulting in an effective decrease of the body effect coefficient by the factor \hat{Q}_B/Q_B . This may make sense intuitively since, in a short-channel device, the substrate's control on the channel is reduced, as expected from Fig. 6.4c, and much of the channel is controlled by the gate, the source, and the drain. However, defining $(\hat{Q}_B/Q_B)\gamma$ as the new body-effect "coefficient" is not very practical, since (\hat{Q}_B/Q_B) is itself a function of V_{SB} , as we will see. From the above equations we can write

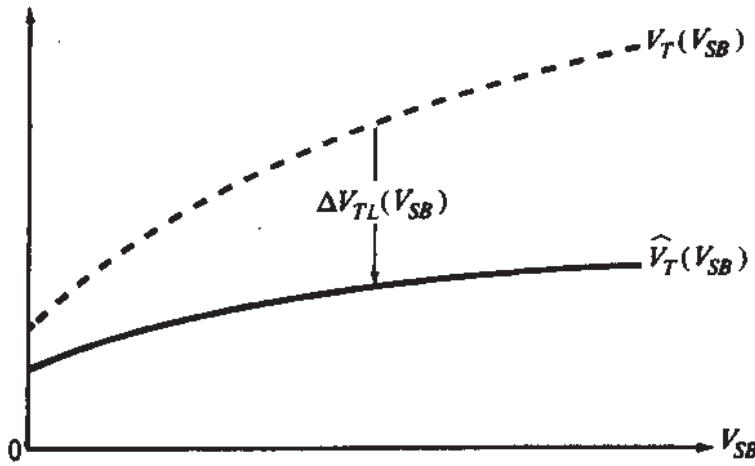
$$\boxed{\hat{V}_T = V_T + \Delta V_{TL}} \quad (6.3.3)$$

where ΔV_{TL} is the "threshold change" due to the short-channel effect, given by

$$\Delta V_{TL} = -\left(1 - \frac{\hat{Q}_B}{Q_B}\right) \gamma \sqrt{\phi_0 + V_{SB}} \quad (6.3.4)$$

With $|\hat{Q}_B| < |Q_B|$, ΔV_{TL} is a *negative* number as expected from the above discussion.

The meaning of ΔV_{TL} is illustrated in Fig. 6.6. Note that the V_{SB} dependence of ΔV_{TL} in (6.3.4) does not include the total effect of V_{SB} on \hat{V}_T . For a given V_{SB} , one must first find V_T for a long-channel device and then change it by the corresponding ΔV_{TL} to arrive at the value of \hat{V}_T . The value of \hat{V}_T obtained in this way is widely used in place of V_T in long-channel equations, such as (4.5.37). There is a large amount of literature on how \hat{V}_T should be calculated, but very little on why it is justifiable to use this value in long-channel equations for the drain current (derived by using the

**FIGURE 6.6**

Threshold voltage of long-channel transistor, V_T ; effective threshold voltage of short-channel transistor with a given channel length, \hat{V}_T ; and difference $\Delta V_{TL} = \hat{V}_T - V_T$, all as a function of V_{SB} .

gradual-channel approximation) and expect the result to be correct for short-channel devices. The issue is not just what value of threshold voltage should be used in such equations when modeling short-channel devices, but rather why the very *form* of such equations should be valid for short-channel devices in the first place. A convincing argument in this regard is not apparent. However, satisfactory agreement with measurement has led to wide use of the aforementioned approach.

Of the many techniques proposed for determining the quantity \hat{Q}_B/Q_B , let us consider one as an example.⁵³ The n^+ -region edge will be assumed cylindrical with radius equal to the junction depth d_j , as shown in Fig. 6.7a. Next, the depletion regions around each of them and below the inversion layer are drawn as if each existed by itself (i.e., no interaction is assumed), and then they are joined together. The width of all three regions is considered equal by assuming that the junction built-in potential ϕ_{bi} is equal to ϕ_0 ; thus, this width is, from (3.4.8),

$$d_B = \zeta \sqrt{\phi_0 + V_{SB}} \quad (6.3.5a)$$

with

$$\zeta = \sqrt{\frac{2\epsilon_s}{qN_A}} \quad (6.3.5b)$$

In this picture, L is assumed to be at least large enough so that a region of trapezoidal cross section can be defined as shown in Fig. 6.7a. \hat{Q}_B is taken to be equal to the charge in that region, and Q_B is assumed to be the charge corresponding to a rectangle of the same depth and length L . Simple geometry then gives (Prob. 6.7):

$$\frac{\hat{Q}_B}{Q_B} = 1 - \frac{d_j}{L} \left(\sqrt{1 + \frac{2d_B}{d_j}} - 1 \right) \quad (6.3.6)$$

\hat{V}_T and ΔV_{TL} can now be found from (6.3.2b) and (6.3.4) to (6.3.6). If L is too small, the trapezoid in Fig. 6.7a becomes a triangle; this case is considered in Prob. 6.8. In the following, we assume that $d_B < L/2$, so that this case does not occur.

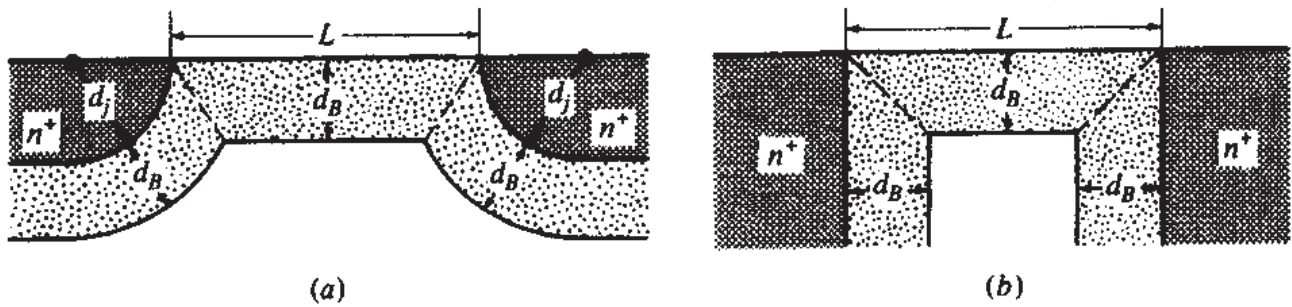


FIGURE 6.7

(a) Trapezoid approximation used for deriving the effective bulk charge; (b) the limiting case of (a) for very deep n^+ regions.

For some practical applications the above formulation is too complicated. It has been suggested that for approximate calculations the quantity in parentheses in (6.3.6) can be approximated by using a Taylor expansion and neglecting the high-order terms,⁵⁹ resulting in (Prob. 6.7):

$$\frac{\hat{Q}_B}{Q_B} \approx 1 - \frac{d_B}{L} \quad (6.3.7)$$

This expansion will be more accurate if d_B/d_j is small. Negligible d_B/d_j corresponds to Fig. 6.7b, for which (6.3.7) can easily be obtained directly. Since, for larger values of d_B/d_j , the expression can be in considerable error, one can help expand its region of validity by introducing a parameter:

$$\frac{\hat{Q}_B}{Q_B} \approx 1 - \beta_1 \frac{d_B}{L} \quad (6.3.8)$$

where nominally $\beta_1 = 1$, but one can allow for empirically adjusting this parameter to obtain best overall fit in a given region.† Using this equation and (6.3.5) in (6.3.2b), we obtain

$$\hat{V}_T \approx V_{FB} + \phi_0 + \gamma \sqrt{\phi_0 + V_{SB}} \left(1 - \frac{\beta_1 \zeta}{L} \sqrt{\phi_0 + V_{SB}} \right)^{\frac{2}{3}} \quad (6.3.9)$$

†The same can be done with ϕ_0 in the equations in this section. Recall that the value of this parameter was somewhat ambiguous even for long-channel devices (see Secs. 2.5.2 and 4.5.1). Thus, although the value of $2\phi_F$ is widely used for ϕ_0 , better results can be obtained if one allows ϕ_0 to be adjusted empirically. Given the heavily empirical nature of the charge-sharing model we are discussing, this is reasonable.

‡It would appear from this equation that, with a sufficiently large V_{SB} , dV_T/dV_{SB} can become negative. However, it can be easily checked by differentiation that, for this to happen, V_{SB} would need to be so large that $d_B > L/2$. In that case, though, the development of this equation is not valid [(see after (6.3.6)]. This is a good example of the wrong conclusions that can be reached by blindly using an equation and forgetting the assumptions that led to its development.

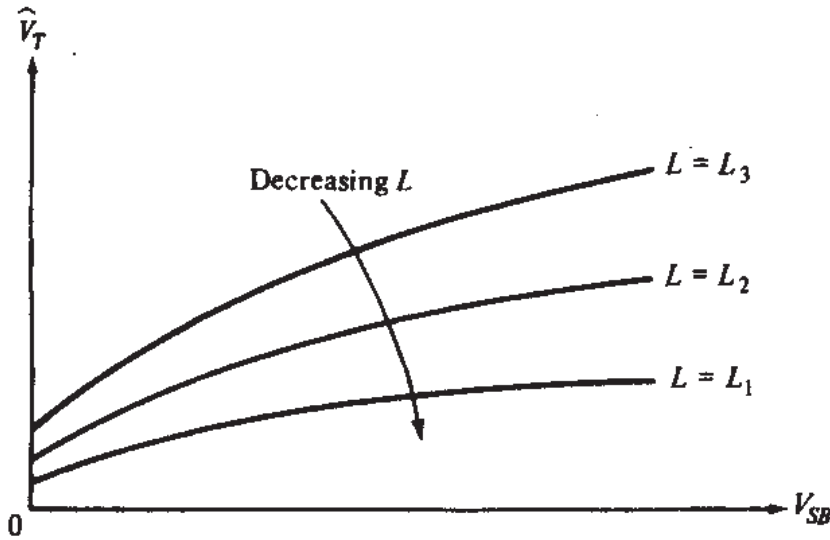


FIGURE 6.8

Effective threshold voltage as a function of source-substrate bias for various channel lengths.

The factor in parentheses can be thought of as an effective reduction in the body effect coefficient. As expected, this reduction is more severe for smaller L . The behavior predicted by the above equation is illustrated in Fig. 6.8. It is seen that, for short channels and large V_{SB} , the dependence of the effective threshold on V_{SB} diminishes. This is because the factor in parentheses in (6.3.9) becomes very small. Physically, this corresponds to the fact that the lower base of the trapezoid in Fig. 6.7 diminishes in length. The bottom of the trapezoid is then practically cut off from the rest of the substrate, and, thus, the control of the substrate on the charge inside the trapezoid is small (Prob. 6.8).

A calculation of the reduction ΔV_T corresponding to the above effects is illuminating. Using (6.3.5) and (6.3.8) in (6.3.4), and recalling that $\gamma = (2q\epsilon_s N_A)^{1/2}/C'_{ox}$, with $C'_{ox} = \epsilon_{ox}/t_{ox}$, we obtain

$$\Delta V_{TL} \approx -2\beta_1 \frac{\epsilon_s}{\epsilon_{ox}} \frac{t_{ox}}{L} (\phi_0 + V_{SB}) \quad (6.3.10)$$

From this equation it is clear that the channel length and the oxide thickness have a competing role. Although decreasing L tends to increase the short-channel effect, decreasing t_{ox} tends to decrease it. This is because then the gate is closer to the channel and is thus better able to keep control of the depletion region charge (as opposed to releasing this control to the other structures surrounding the channel).

Good agreement has been reported between (6.3.9) and (6.3.10) and experimental results, with $\beta_1 = 1$.⁵⁹ However, other work has produced models equivalent to taking β_1 proportional to a negative power of L ,⁶⁷ or even replacing β_1/L by a negative exponential in L .^{68,73}

EFFECT OF DRAIN-SOURCE VOLTAGE. The above results have been derived for negligible V_{DS} . For fixed V_{SB} , if V_{DS} (and thus V_{DB}) is increased, the depletion region

width around the drain will widen. This, in turn, following the lines of the above arguments, will decrease $|\hat{Q}_B|$ below the values obtained so far, thus further decreasing \hat{V}_T . Thus, for short-channel devices, \hat{V}_T becomes a decreasing function of V_{DS} ! The effective charge \hat{Q}_B is still calculated using a method similar to the one illustrated in Fig. 6.7a, only now the trapezoid there will become distorted. The resulting expression is considered in Prob. 6.11. If d_j is large and certain simplifications are made (Prob. 6.11), we can obtain

$$\frac{\hat{Q}_B}{Q_B} \approx 1 - \beta_1 \frac{1}{L} \frac{d_{BS} + d_{BD}}{2} \quad (6.3.11)$$

where d_{BS} and d_{BD} are the depletion region widths of the source and the drain, respectively, and β_1 is again a constant, nominally equal to 1 but which can be adjusted for better fit to experimental results. Note that for $V_{DS} = 0$ ($d_{BS} = d_{BD} = d_B$), the above equation reduces to (6.3.8). The quantity $(d_{BS} + d_{BD})/2$ is, from (6.3.5a) and a similar equation for the drain depletion region,

$$\frac{d_{BS} + d_{BD}}{2} \approx \frac{\zeta}{2} \left(\sqrt{\phi_0 + V_{SB}} + \sqrt{\phi_0 + V_{DB}} \right) \quad (6.3.12)$$

We can write $V_{DB} = V_{SB} + V_{DS}$. For small V_{DS} , the second square root can be expanded in a Taylor series around $V_{DS} = 0$ and approximated by the first two terms; this results in (Prob. 6.11):

$$\frac{d_{BS} + d_{BD}}{2} \approx \zeta \left(\sqrt{\phi_0 + V_{SB}} + \frac{\beta_2 V_{DS}}{\sqrt{\phi_0 + V_{SB}}} \right) \quad (6.3.13)$$

where $\beta_2 = 0.25$ results from the expansion. To increase the region of validity of this expression, however, empirical adjustment of β_2 may be needed. Using now (6.3.13), (6.3.11), and (6.3.2b), we obtain,

$$\hat{V}_T \approx V_{FB} + \phi_0 + \gamma \sqrt{\phi_0 + V_{SB}} \left[1 - \frac{\beta_1 \zeta}{L} \left(\sqrt{\phi_0 + V_{SB}} + \frac{\beta_2 V_{DS}}{\sqrt{\phi_0 + V_{SB}}} \right) \right] \quad (6.3.14)$$

and thus

$$\Delta V_{TL} \approx -2\beta_1 \frac{\epsilon_s}{\epsilon_{ox}} \frac{t_{ox}}{L} \left[(\phi_0 + V_{SB}) + \beta_2 V_{DS} \right] \quad (6.3.15)$$

The behavior predicted by the above equations is shown in Fig. 6.9.

Although (6.3.15) was derived by using the charge-sharing idea, it predicts a variation of V_T with V_{DS} in the same direction as considerations using the barrier-lowering concept^{60,66} discussed in the beginning of this section. The name *drain-*

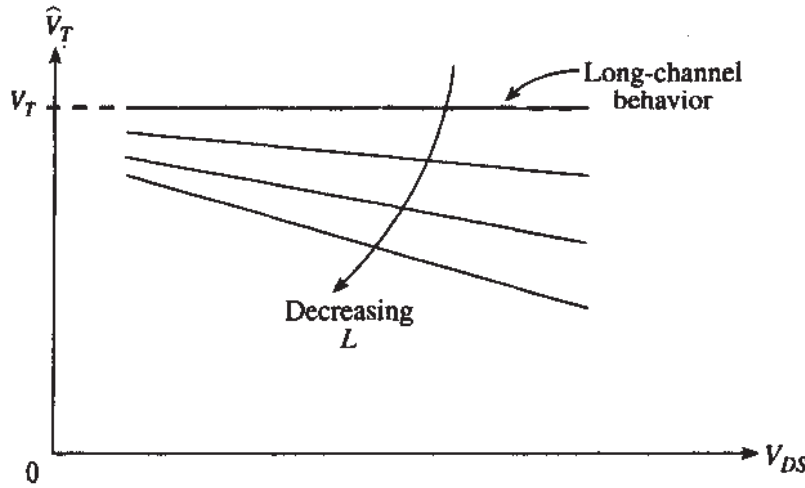


FIGURE 6.9
Effective threshold voltage vs. V_{DS} , with L as a parameter.

induced barrier lowering (DIBL) is used to describe the effect of the drain on the barrier.⁶⁶ Although the above expressions were based on the assumption of small V_{DS} , in practice it is sometimes found that they are reasonable even for relatively large V_{DS} ranges, provided an appropriate value is chosen for β_2 . This value will be different for different L ; in fact, for some devices it has been found that β_2 is roughly proportional to $1/L$.³⁷ To explain such behavior, one should consider a more detailed picture in which the field lines emanating from the drain not only affect the channel charge near the drain but rather affect it throughout the channel length. Two-dimensional numerical simulations are needed to get quantitative results in this case.⁷⁷

It is important to note that the effect of the drain on the channel depletion region charge, resulting in (6.3.15), continues even beyond pinchoff. Thus, even if channel length modulation is neglected, I_{DS} in the “saturation” region will not saturate. It will continue to increase as V_{DS} is increased, since \hat{V}_T will keep decreasing. Note, also, that a device which has been turned off by reducing V_{GS} sufficiently below V_T , may turn on just by increasing V_{DS} if \hat{V}_T is lowered enough by the above-mentioned effect. This can have a serious effect in some digital applications if not properly considered.

The onsets of moderate and of strong inversion, V_M and V_H , stay close to V_T and follow qualitatively a similar dependence on L , V_{SB} , and V_{DS} . From what we have seen so far, to a first order, short-channel effects can be thought of as a shift in the I_{DS} - V_{GS} curves to lower V_{GS} values, with the shift being approximately equal to ΔV_{TL} .

Let us now briefly look at the current in weak inversion. Assuming that (4.6.13) holds, but with V'_M shifted by ΔV_{TL} , we see that

$$I_{DS, \text{ short channel}} = I_{DS, \text{ long channel}} e^{-\Delta V_{TL}/(n\phi_t)}, \quad \text{weak inversion} \quad (6.3.16)$$

Again, since ΔV_{TL} depends on V_{DS} , I_{DS} will depend on V_{DS} for any value of the latter; in other words, no real “saturation” of I_{DS} will be observed. This is verified by experimental results.^{38,61,64,66}

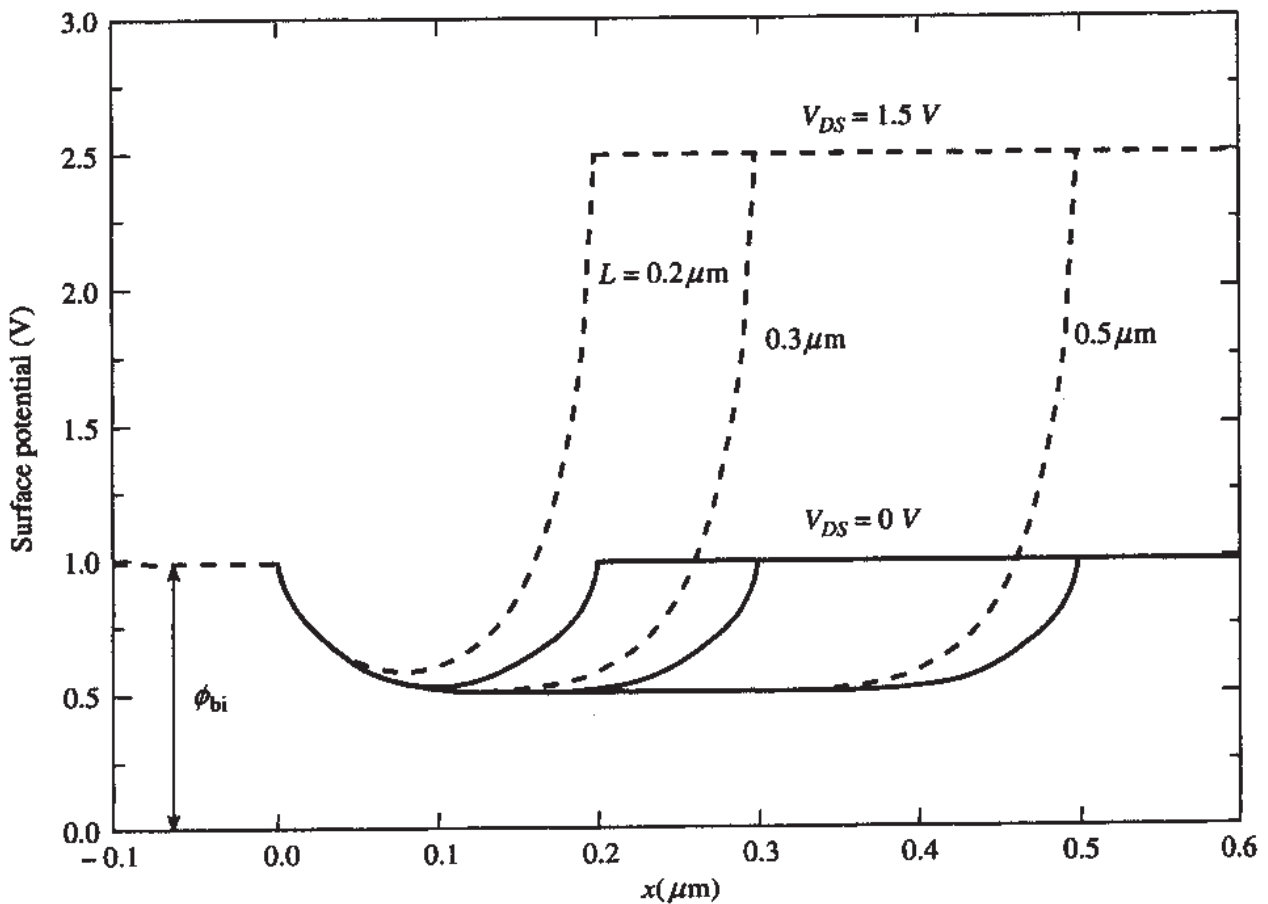


FIGURE 6.10

Plot of surface potential vs. length along the channel for three nMOSFETs with $L = 0.2, 0.3$, and $0.5 \mu\text{m}$, in weak inversion with $V_{GS} = 0 \text{ V}$ and $V_{DS} = 0 \text{ V}$ (solid lines) and 1.5 V (dashed lines).

RESULTS FROM 2-D AND PSEUDO-2-D ANALYSES. Figure 6.10 illustrates the effects of L and V_{DS} on the surface potential. Although these results are obtained by two-dimensional (2-D) numerical simulation,⁸⁹ they should make sense qualitatively on the basis of our discussion so far. The channel and source/drain doping concentrations for the three channel lengths is the same and designed such that the $L = 0.5 \mu\text{m}$ device operates as a “long-channel” device while the $L = 0.2 \mu\text{m}$ device exhibits clear short-channel effects. Note that the potential minimum for the $0.2 \mu\text{m}$ device even at $V_{DS} = 0 \text{ V}$ is higher compared to that of both 0.3 and $0.5 \mu\text{m}$ devices. This corresponds to an energy barrier lowering, as explained earlier, which results in a V_T decrease. Note now what happens as V_{DS} is raised to 1.5 V . The potential minimum for the $0.2\text{-}\mu\text{m}$ device is further raised while that of the $0.5\text{-}\mu\text{m}$ device is unaffected and for the $0.3\text{-}\mu\text{m}$ device is just barely raised. This is a manifestation of the DIBL effect discussed above. So, clearly the $0.2\text{-}\mu\text{m}$ device is operating in the short-channel regime, the $0.3\text{-}\mu\text{m}$ device is in the borderline between short and long regime, and the $0.5\text{-}\mu\text{m}$ device is clearly in the long-channel regime.

One should note here that in Fig. 6.10 the surface potential at $V_{GS} = 0 \text{ V}$, which corresponds to the weak inversion regime for these devices, is never flat between source and drain, even for the device that does not exhibit short-channel effects. It would appear that the form of the inversion current equation (4.6.13), which is

derived on the assumption of flat surface potential, may not be valid. It turns out that while some modification to the equation may be required,⁹⁰ the equation can still be used and the channel-length and DIBL effects are captured by the modification of V_T , as in (6.3.16).

While it is not possible to obtain the accuracy of 2-D numerical simulations in an analytical expression, several quasi-2-D solutions to Poisson's equation have been proposed in the literature.^{31,90-97} They all are based on the idea that the barrier lowering we just described can be linked to ΔV_{TL} . A representative result for ΔV_{TL} from such solutions, the details of which will be omitted here, is⁹⁵

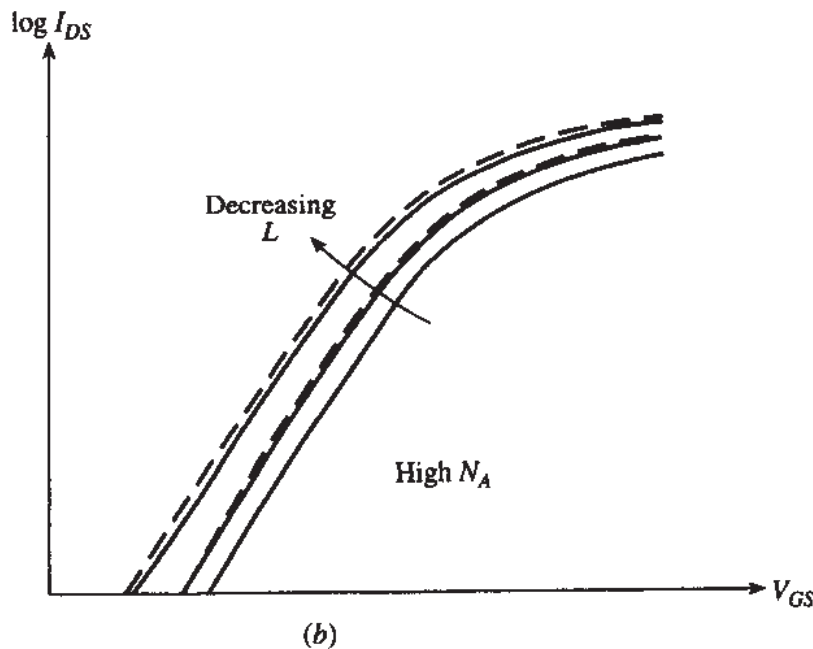
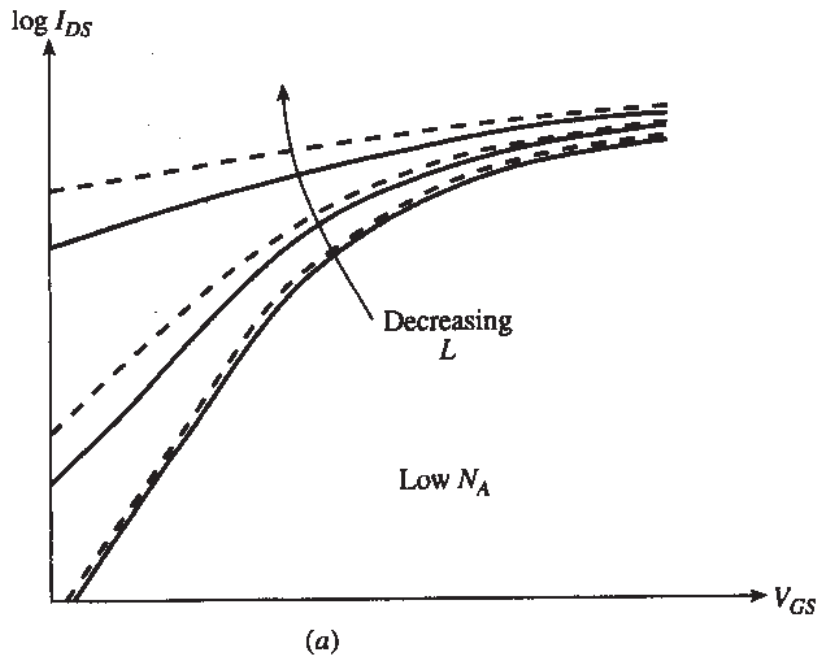
$$\Delta V_{TL} \approx -[3(\phi_{bi} - \phi_0) + V_{DS}]e^{-L/\lambda} \quad (6.3.17)$$

where ϕ_{bi} is the source (or drain) to channel junction built-in potential, defined in (1.5.1), and λ is a characteristic length, or "gauge" for the channel length given by

$$\lambda = \sqrt{\frac{\epsilon_s t_{ox} d_B}{\epsilon_{ox} \beta_3}} \quad (6.3.18)$$

with d_B being the depth of the depletion region below the channel, defined in (6.3.5a), and β_3 a fitting parameter close to 1. Equation (6.3.17) is derived under the assumption that $L \gg d_B$ ($L \geq 4d_B$ is sufficient). These results predict again the general behavior shown in Fig. 6.9.

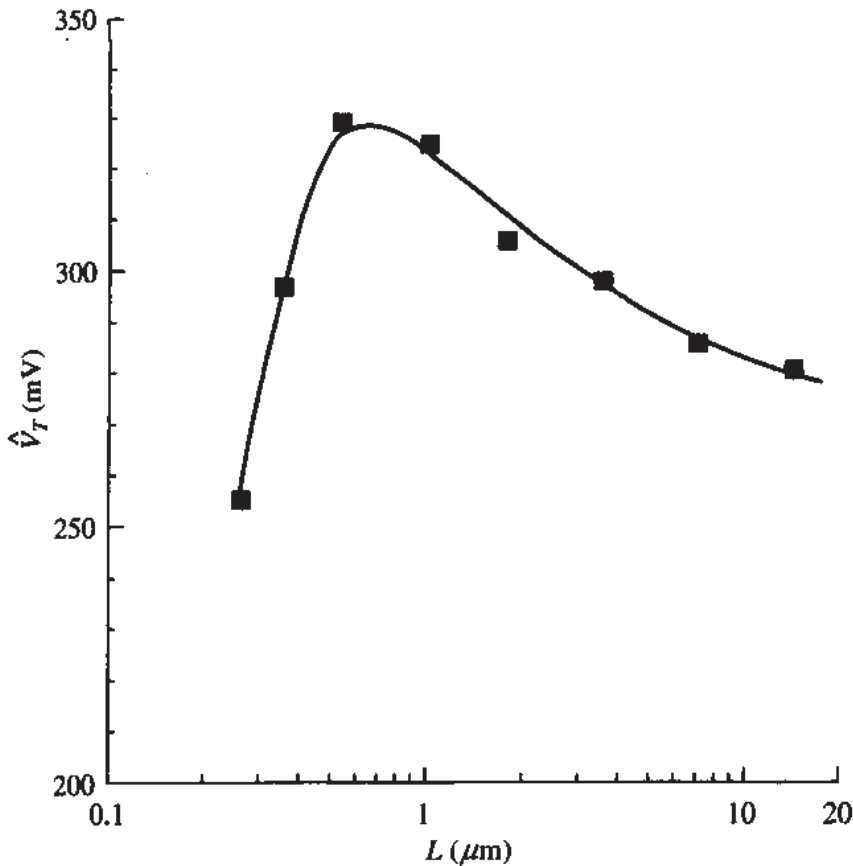
Equation (6.3.17) has two qualitative advantages over (6.3.15). First, it predicts that $\Delta V_{TL} \propto e^{-L/\lambda}$, a fact that is confirmed experimentally in any given technology, i.e., for a given λ . Note that (6.3.15) predicts a $1/L$ dependence of ΔV_{TL} which, for small L , is not as steep as $e^{-L/\lambda}$. Second, (6.3.17) with (6.3.18) predicts that $|\Delta V_{TL}|$ decreases as the channel doping increases, because then d_B and therefore λ decrease, while (6.3.15) has no doping dependence. This doping dependence of ΔV_{TL} is a well-known fact in device engineering. Finally, (6.3.17) predicts an increase in DIBL with increasing V_{SB} because d_B and therefore λ increase, while (6.3.15) predicts no dependence of DIBL on V_{SB} (β_2 in that equation is nominally a constant). Again, this DIBL increase with V_{SB} is a well-known fact. On the other hand, it is well-known that $|\Delta V_{TL}|$ decreases with decreasing source/drain junction depth d_j , yet neither (6.3.15) nor (6.3.17) predicts a dependence on d_j , although it has been suggested that (6.3.18) may be modified empirically to include it.⁹⁵ The above discussion begs the question, "Which model, charge-sharing or quasi-2-D Poisson solution, should be used?" It turns out that the charge-sharing model is preferred for compact MOSFET models used in circuit simulation programs (e.g., Ref. 98), because it gives simpler equations that are easily adapted to nonuniform (i.e., implanted) channels (see, e.g., Ref. 88). This model is reasonably accurate provided doping and junction depth are within a narrow range, as is the case for any given technology generation. On the other hand, for compact models used to predict technology trends, such as scaling (Sec. 6.7), where junction depths and channel doping vary over a broad range, the quasi-2-D Poisson solution model is preferable. The two modeling approaches are sometimes mixed.

**FIGURE 6.11**

Log I_{DS} vs. V_{GS} for three different channel lengths (modified from Ref. 86). Solid lines are for one value of V_{DS} ; broken lines are for another, larger V_{DS} value. $V_{SB} = 0$. (a) Low N_A , (b) high N_A .

The concept of simple shift of the I_{DS} - V_{GS} (and $\log I_{DS}$ - V_{GS}) characteristics by $\Delta V_{TL}(L, V_{BS}, V_{DS})$ is a very convenient way to capture the short-channel effects. However, as the channel becomes very short, for the choice of physical device parameters such as t_{ox} , N_A , and d_j (see discussion on MOSFET scaling in Sec. 6.7), this simple concept fails. Most noticeable is the failure of the $\log I_{DS}$ vs. V_{GS} characteristics to maintain their slope if L is too small and V_{DS} is increased.

In such cases, the slope of $\log I_{DS}$ becomes very small, and the device cannot be turned off adequately even if V_{GS} is decreased significantly. This behavior is predicted by two-dimensional simulations (see also the discussion of the “punch-through” effect in Sec. 6.4). An example⁸⁶ is shown in Fig. 6.11a. One might expect that it should be possible to eliminate the undesirable behavior for very short channels by increasing N_A , since then the source and drain depletion region widths will

**FIGURE 6.12**

Plot of \hat{V}_T for a set of nMOSFET devices exhibiting the *reverse* short-channel effect. This effect is manifested by the slow rise of V_T as L decreases from $12\ \mu\text{m}$ to about $0.5\ \mu\text{m}$. The rapid decrease of V_T below $0.5\ \mu\text{m}$ is due to the “normal” short-channel effect discussed in Sec. 6.3.2.

decrease, as predicted by (6.3.5), or λ will decrease as predicted by (6.3.18) and (6.3.5). Indeed, raising N_A leads to the results in Fig. 6.11b. Here it is also interesting to note that for the longest channel the current does not depend on V_{DS} , just as is expected from long-channel theory for weak inversion and $V_{DS} > 3\phi_t$. However, a V_{DS} dependence is apparent for the shorter channel lengths in the figure, as expected from the above discussion.

With considerable computational complexity, it is possible to extend the charge sheet model of Sec. 4.3.1 to cover short-channel effects.^{46,74,80,88,97}

THE REVERSE SHORT-CHANNEL EFFECT. On the basis of the charge-sharing or barrier-lowering theories just discussed, we saw that \hat{V}_T is expected to decrease monotonically with decreasing L . However, it is often observed that \hat{V}_T first increases with decreasing L , for moderately small L , before it decreases as predicted from theory. Figure 6.12 depicts the “hump” effect on \hat{V}_T for a typical technology.

Several theories have been proposed to explain this behavior, which has been called the *reverse short-channel effect* (RSCE). In all cases some form of nonuniformity along the channel length is invoked that would result in nonuniform threshold voltage along the channel. For example Q'_o or N_A nonuniformities have been proposed.^{99,100} The current thinking at the time of the writing of this book is that the latter is more likely. The physical reasons for this nonuniformity, which can itself be a function of gate length, fall outside the scope of this book. Interested readers can pursue this further in the cited literature. Generally, attempts are made to minimize this effect but it is commonly observed at varying magnitudes in modern short-channel technologies.

6.3.3 Narrow-Channel Devices

A cross-section of a device's channel along its *width* is shown in Fig. 6.13a. This figure is highly idealized for simplicity. Real devices, for most fabrication processes, look as in Fig. 6.13b or c. A thick oxide, called *field* oxide, becomes thinner and defines the channel width as equal to the extent of the thin oxide region.

Figure 6.13b depicts the so-called LOCOS (*local oxidation of silicon*) isolation¹⁰¹ which results in the gradual transition from thick (field) oxide to thin (gate) oxide in what is called the “bird’s beak” region. Figure 6.13c depicts a “fully recessed” isolation structure, also called *shallow-trench isolation* (STI), which is being used^{101,102} in the sub-0.35 μm CMOS generations. We will discuss each case separately.

LOCOS ISOLATION. In Fig. 6.13a or b, the depletion region is not limited to just the area below the thin oxide. This is because some of the field lines emanating from the gate charges terminate on ionized acceptor atoms on the sides; this is shown in Fig. 6.14a and b. These lines constitute what is called a *fringing field*. If W is large, the part of the depletion region on the sides is a small percentage of the total depletion region volume and can be neglected. However, with small W values, the side parts become a large percentage of the total. In contrast to the short-channel case, now the gate is responsible for depleting a region which is *larger* than what our theory in Chap. 4 would predict^{54–56,59,62,70,74,76–81,83,85–87} (a long channel is assumed in this discussion). Thus, it takes a higher V_{GS} value to deplete that amount before an inversion layer can be formed, and V_M , V_H , and V_T must be replaced by *larger*, “effective quantities.” In addition, an increase in V_{SB} , while keeping V_{GS} constant, increases $|Q_B|$ both under the gate and in the side parts, and, thus, the body effect will be more pronounced than in the wide-channel case. Assuming again $V_{DS} = 0$ and using an empirical approach similar to that in the previous subsection, an effective depletion region charge \hat{Q}_{B1} can be employed. This will result in an effective threshold \hat{V}_T given by

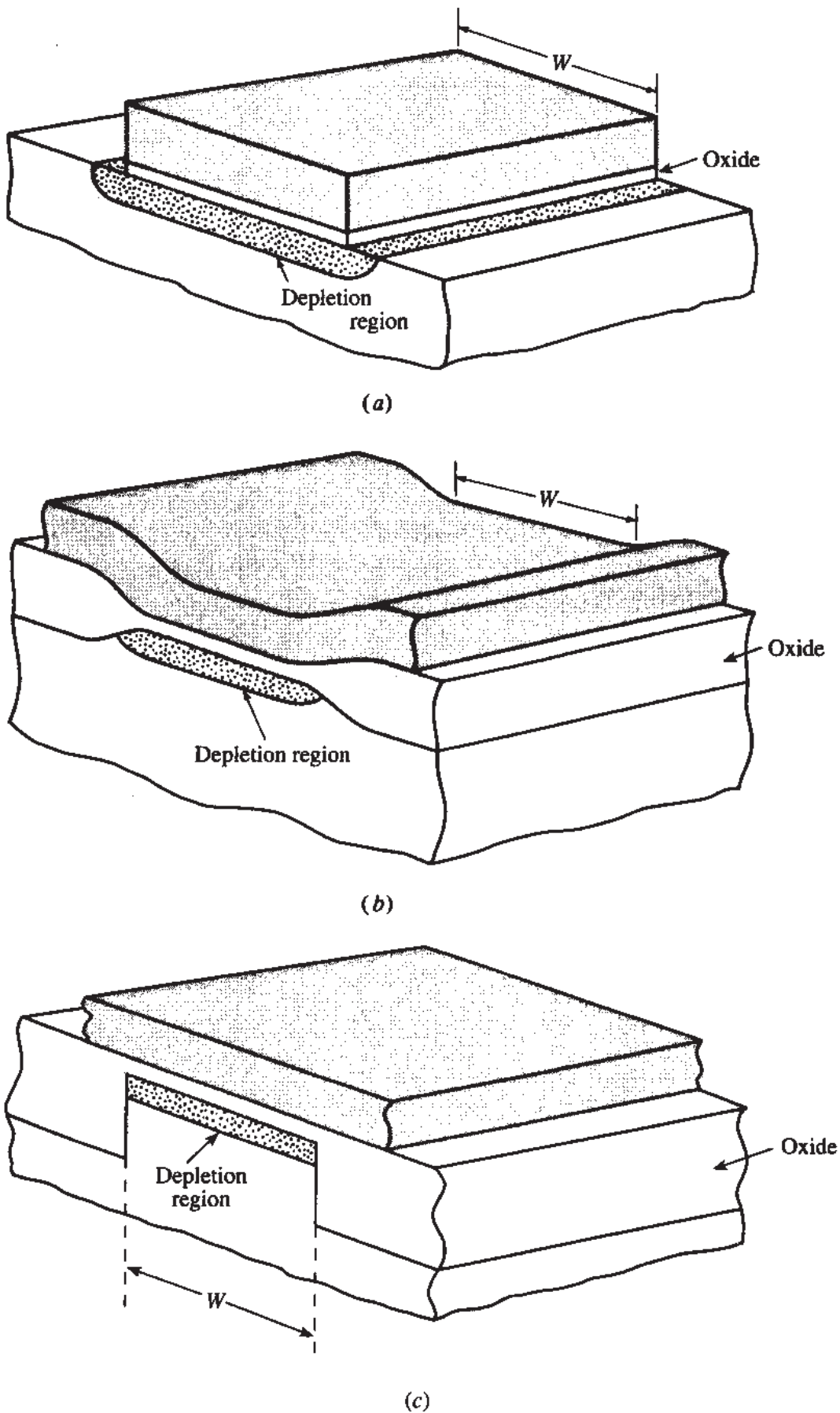
$$\hat{V}_T = V_{FB} + \phi_0 + \frac{\hat{Q}_{B1}}{Q_B} \gamma \sqrt{\phi_0 + V_{SB}} \quad (6.3.19)$$

which corresponds to (6.3.2b), only here \hat{Q}_{B1}/Q_B is larger than unity. Thus, compared to the long- and wide-channel threshold V_T given by (6.3.1b), we have here an effective threshold:

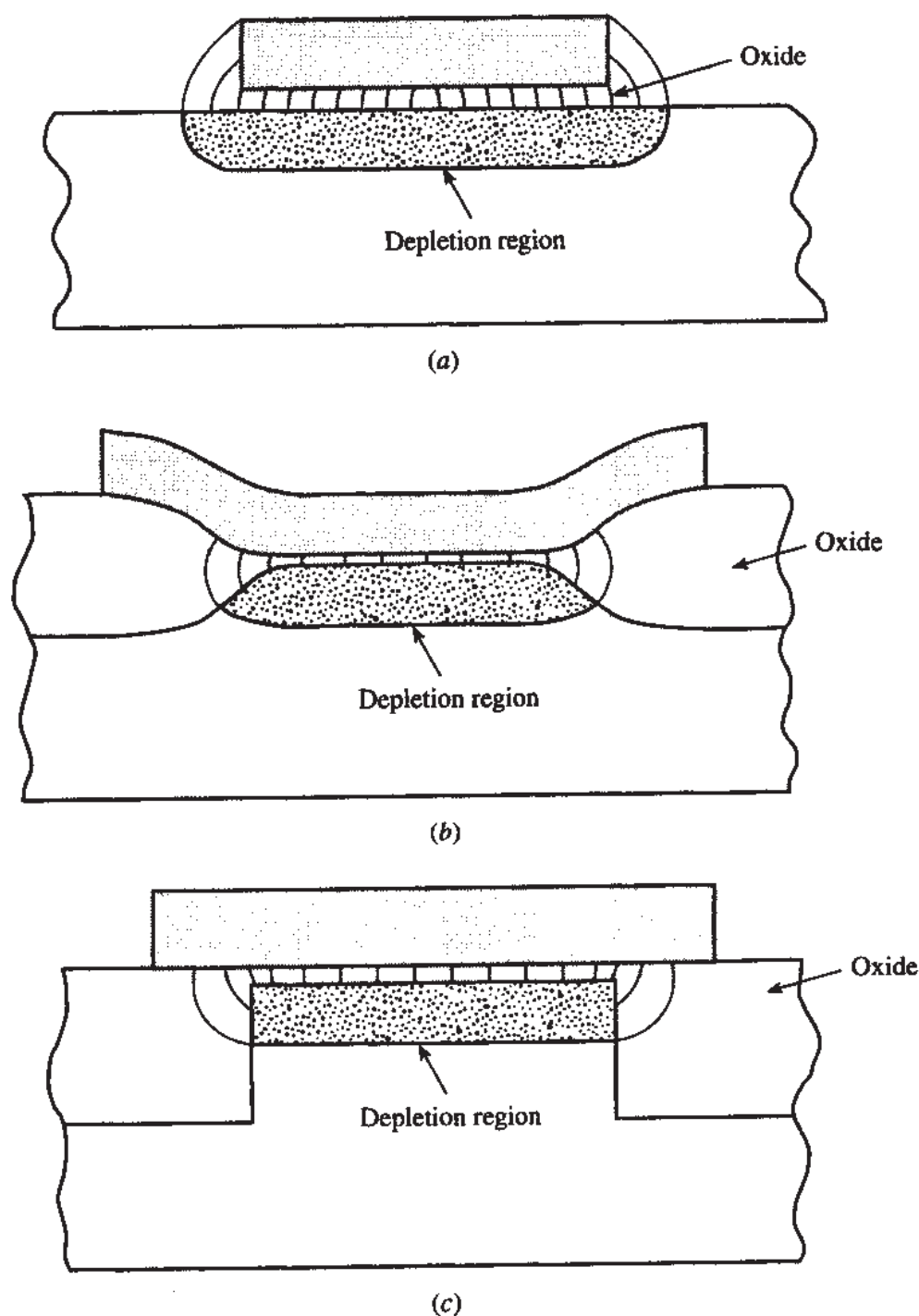
$$\hat{V}_T = V_T + \Delta V_{TW} \quad (6.3.20)$$

with

$$\Delta V_{TW} = \left(\frac{\hat{Q}_{B1}}{Q_B} - 1 \right) \gamma \sqrt{\phi_0 + V_{SB}} \quad (6.3.21)$$

**FIGURE 6.13**

Cross sections along the width of MOSFET transistors. (a) Simplified structure, where the device width is defined by the width of the gate; (b) more realistic representation of a LOCOS-isolated device with "bird's beak" in the transition from thick field oxide to the channel oxide; (c) realistic representation of a shallow-trench-isolated (STI) device.

**FIGURE 6.14**

Schematic cross-sections along the width of MOSFET transistors depicting the effect of gate fringing electric lines on the extent of the depletion region. (a) Simplified structure; (b) LOCOS-isolated MOSFET; (c) shallow-trench-isolated (STI) MOSFET.

One of the ways suggested for the empirical evaluation of \hat{Q}_{B1}/Q_B uses the simplified picture of Fig. 6.13a and assumes that the side parts of the depletion region have a quarter-circle cross-section.⁵⁹ This gives, assuming negligibly small V_{DS} (Prob. 6.13),

$$\frac{\hat{Q}_{B1}}{Q_B} \approx 1 + \beta_4 \frac{\pi}{2} \frac{d_B}{W} \quad (6.3.22)$$

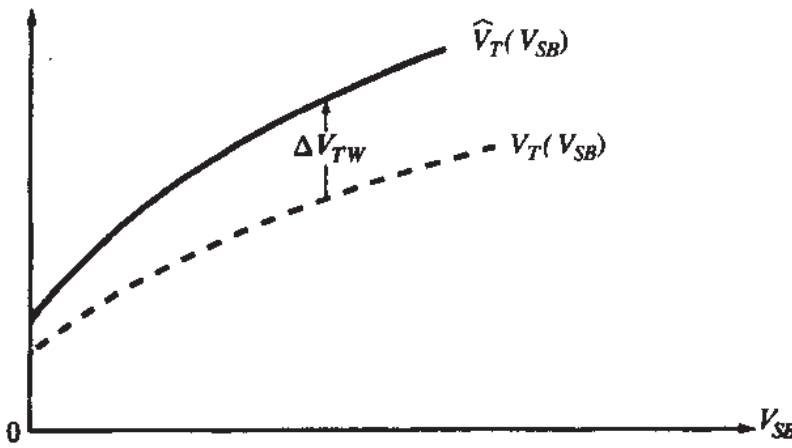


FIGURE 6.15

Threshold voltage of a wide transistor, V_T ; effective threshold voltage of narrow-channel LOCOS-isolated transistor with a given channel width, \hat{V}_T ; and difference $\Delta V_{TW} = \hat{V}_T - V_T$, all as a function of V_{SB} .

where d_B is the depth of the depletion region given by (6.3.5), and $\beta_4 = 1$ nominally. As usual, β_4 can be empirically adjusted for best fit. Comparing the simplified and "real" pictures in Fig. 6.14a and b, it is rather obvious that such an empirical adjustment is appropriate. Using (6.3.22) and (6.3.5) in (6.3.19) we obtain the corresponding effective threshold \hat{V}_T :

$$\hat{V}_T \approx V_{FB} + \phi_0 + \gamma \sqrt{\phi_0 + V_{SB}} \left(1 + \frac{\beta_4 \zeta \pi}{2W} \sqrt{\phi_0 + V_{SB}} \right) \quad (6.3.23)$$

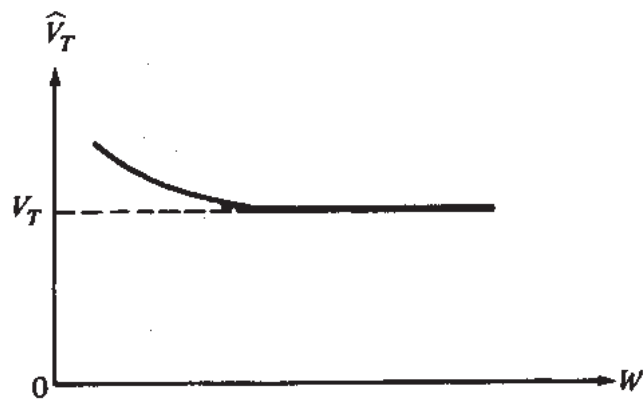
and, using the above and (6.3.1b) in (6.3.20), with $\gamma = (2q\epsilon_s N_A)^{1/2}/C'_{ox}$ and $C'_{ox} = \epsilon_{ox}/t_{ox}$,

$$\Delta V_{TW} \approx \beta_4 \pi \frac{\epsilon_s}{\epsilon_{ox}} \frac{t_{ox}}{W} (\phi_0 + V_{SB}) \quad (6.3.24)$$

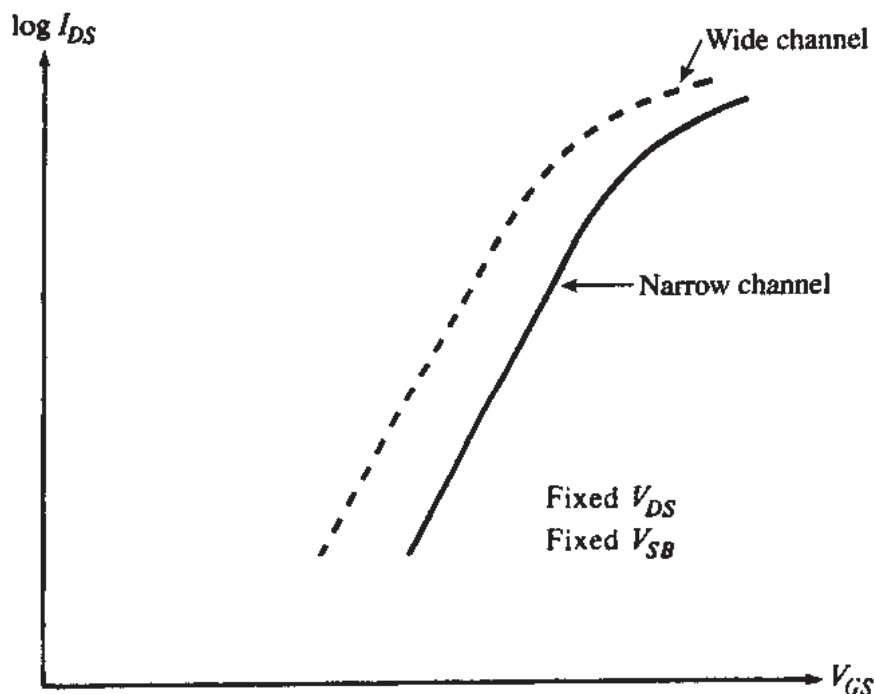
This equation is in agreement with our intuitive picture above. Values of 50 to 200 mV are typical for ΔV_{TW} for minimum channel widths. The quantity ΔV_{TW} does not include the total dependence of \hat{V}_T on V_{SB} . V_T must be first found for the wide-channel case for a known V_{SB} and then augmented by ΔV_{TW} . Note that ΔV_{TW} in the narrow-channel LOCOS-isolated device is positive, whereas the quantity ΔV_{TL} used in the discussion of short-channel devices was negative (assuming no reverse short-channel effects are present). The narrow-channel effect for LOCOS-isolated devices is illustrated in Figs. 6.15, 6.16, and 6.17.

The value of \hat{V}_T for devices with small W (but large L) is commonly assumed independent of V_{DS} . [The above arguments would tend to predict an increase with V_{DS} (Prob. 6.14)].

Depending on fabrication details, the above analysis may give satisfactory results if W is not too small. For some devices, though, it is observed that, in order to fit measured results, it is not sufficient to replace V_T by the effective threshold \hat{V}_T as

**FIGURE 6.16**

Effective threshold voltage vs. channel width (assuming LOCOS fabrication process).

**FIGURE 6.17**

Log I_{DS} vs. V_{GS} for fixed V_{DS} and V_{SB} , for wide- and narrow-channel devices, assuming LOCOS isolation.

above, and one must, in addition, replace W by an “effective channel width” \hat{W} . Unfortunately, to evaluate \hat{W} one must resort to either measurements or to two-dimensional simulations, including the details of the so-called “channel-stop” regions adjacent to the channel at its two sides. Qualitatively, one observes two effects: (1) \hat{W} tends to increase with increasing gate voltage, since then a wider part of the channel tends to be inverted in Fig. 6.13b. (2) \hat{W} tends to decrease with V_{SB} .⁸⁷ This is because, as two-dimensional simulations show (including the channel-stop regions), the bottom of the depletion region is actually rounded, and the electric field lines in the semiconductor are perpendicular to the surface only near the center of the channel in Figs. 6.13b and 6.14b; toward the sides, they curve sideways. In such cases, the inversion layer on

both sides of the center turns out to be lighter than expected or it practically disappears altogether there. This effect becomes stronger with increasing V_{SB} , causing a reduction in \hat{W} , which can be very significant for some fabrication processes.

As already discussed, the difference between \hat{Q}_{B1} and Q_B arises from the channel edge fringing fields. Another expression suggested for estimating \hat{Q}_{B1}/Q_B for a given $V_{GS} \geq V_T$ (assuming $Q'_I \ll Q'_B$) is⁸⁸

$$\frac{\hat{Q}_{B1}}{Q_B} = \frac{\hat{C}_{GB}}{C'_{ox}WL} \quad (6.3.25)$$

where $C'_{ox}WL$ and \hat{C}_{GB} are the gate-to-channel capacitances needed to support the “ideal” charge Q_B and the “real” charge \hat{Q}_{B1} , respectively. The above expression has been used in the literature to calculate ΔV_{TW} for various isolation configurations.⁸⁸ In the cases shown in Fig. 6.14a and b, \hat{C}_{GB} must include the fringing capacitance $2C_F$ (assuming C_F per side) that supports the extra charge outside the channel. Thus

$$\frac{\hat{Q}_{B1}}{Q_B} = \frac{C'_{ox}WL + 2C_F}{C'_{ox}WL} > 1 \quad (6.3.26)$$

This ratio can be calculated from detailed knowledge of the oxide tapering.

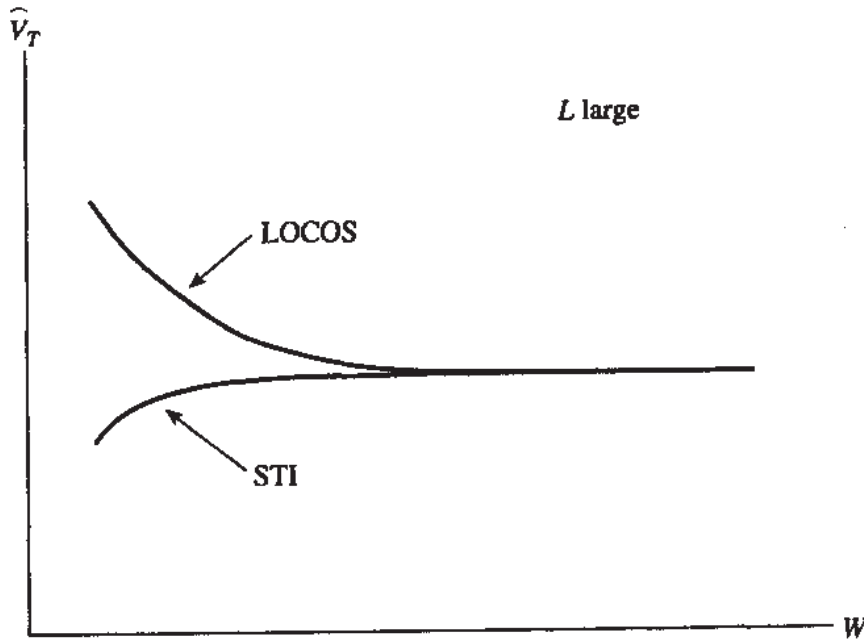
STI ISOLATION-INVERSE NARROW-CHANNEL EFFECT. For the shallow-trench isolation device shown in Figs. 6.13c and 6.14c, the fringing field from the gate regions beyond the channel edges helps support depletion charges *in* the channel. In contrast to the case discussed above, here the fringing field is *useful*, in the sense that it helps make the depletion region deeper, thus increasing the surface potential and helping start the inversion layer. Thus, it now takes a *lower* V_{GS} to deplete the channel before an inversion layer can be formed, and V_M , V_H , and V_T must be replaced by *smaller*, “effective” quantities.

For a uniform, idealized device, V_T is given by (6.3.1a); using total charge and capacitance, we see the last term is equal to $-Q_B/(C'_{ox}WL)$. For the case of Fig. 6.14c, we have instead a *larger* effective capacitance, $C'_{ox}WL + 2C_F$, where C_F is the fringing capacitance on each side. Then the effective threshold \hat{V}_T is assumed to be

$$\hat{V}_T = V_{FB} + \phi_0 - \frac{Q_B}{C'_{ox}WL + 2C_F} \quad (6.3.27)$$

Again, the assumption will be made that the device operates as one with uniform depletion region and capacitance $C'_{ox}WL$, but with smaller “effective” depletion region charge \hat{Q}_{B1} . Then \hat{V}_T can be written

$$\hat{V}_T = V_{FB} + \phi_0 - \frac{\hat{Q}_{B1}}{C'_{ox}WL} \quad (6.3.28)$$

**FIGURE 6.18**

Effective threshold voltage vs. width for identically processed MOSFETs except for the isolation process. As discussed in the text, STI tends to produce a decrease of threshold voltage with decreasing W . Typically, the range of widths over which threshold voltage is decreased in STI is smaller than the range over which threshold is increased in LOCOS isolation.

From the above two equations we have

$$\frac{\hat{Q}_{B1}}{Q_B} = \frac{C'_{ox} WL}{C'_{ox} WL + 2C_F} < 1 \quad (6.3.29)$$

It is then easy to see that (6.3.27) can be written as in (6.3.19), only now \hat{Q}_{B1}/Q_B will be given by (6.3.29). The fringing capacitance C_F can be shown to be¹⁰³⁻¹⁰⁶

$$C_F = \frac{2\epsilon_{ox} L}{\pi} \ln \left(\frac{2t_{Fox}}{t_{ox}} \right) \quad (6.3.30)$$

where t_{Fox} is the thickness of the field oxide. Using (6.3.30) in (6.3.29) we get

$$\frac{\hat{Q}_{B1}}{Q_B} = \frac{W}{W + F} \quad (6.3.31)$$

where

$$F = \frac{4t_{ox}}{\pi} \ln \frac{2t_{Fox}}{t_{ox}} \quad (6.3.32)$$

Note that, in the case of STI, as W decreases \hat{Q}_{B1}/Q_B decreases, just as was the case with the variation of \hat{Q}_B/Q_B with length, as we saw earlier. While the factor F is given explicitly in terms of physical parameters in (6.3.32), typically it is treated as

a fitting parameter in practice. Figure 6.18 compares the W effect in LOCOS- and shallow-trench-isolated MOSFETs. Evidently the factor F for the STI device is quite small in this example.

6.3.4 Summary and Comments

We have seen that several factors affect the value of the effective threshold voltage. From our general expressions for long-channel devices as well as from our discussion of short channels and narrow channels above, we can summarize qualitatively as follows:

The effective threshold voltage decreases when:

1. The substrate doping decreases.
2. The oxide thickness decreases.
3. The channel length decreases (assuming no reverse-short-channel-effect).
4. The junction depth increases.
5. The channel width increases or decreases depending on isolation technology (LOCOS or STI, respectively).

The concept of effective threshold voltage represents an attempt to maintain the form of well-known equations (derived for long- and wide-channel devices) in cases where such equations are, in principle, not valid. In this section, we have presented some representative examples of this approach taken from the literature. As explained, in this approach it has not been possible to avoid some arbitrary assumptions (or, at least, assumptions that are not carefully justified). This is typical of empirical models. Although this has often provided practical results, at times too much attention is paid to deriving elaborate models based on some arbitrary assumptions, rather than taking care to consider the validity of those assumptions. The resulting models, although they may be based on inadequate assumptions, may still be able to model practical devices owing to the large number of parameters they contain, which are “adjusted” empirically. The reader is referred to Chap. 10 for a discussion of the potential drawbacks of such approaches.

6.4 PUNCHTHROUGH

Up to now we have considered electrostatic effects that become important as the channel length decreases, but that are small enough so that they do not compromise the device behavior very much. Perhaps the most important criterion of such “electrostatic integrity” at a given small L is that the weak inversion slope of $\log I_{DS}$ vs. V_{GS} be nearly constant with V_{DS} and almost equal to that of a long-channel device, built with the same process technology. The qualifiers “nearly” and “almost” are used here on purpose because it is common that at any given MOSFET generation, the device with the minimum acceptable channel length at the highest expected V_{DS} will generally show subthreshold-slope decreases of a few percent relative to its long-channel equivalent. Subthreshold slope is usually indicated by the S factor, defined in Sec. 4.6 as $S = dV_{GS}/d \log I_{DS}$. This is numerically equal to the swing of V_{GS} needed

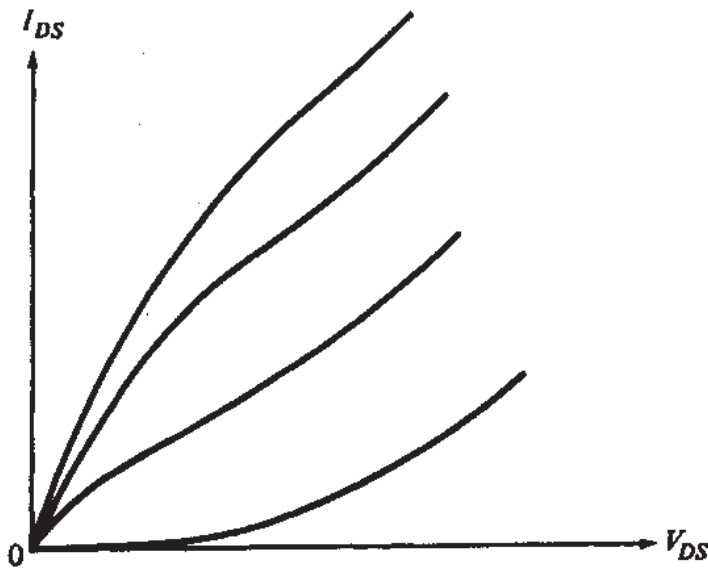


FIGURE 6.19

Typical characteristics of a device with punchthrough problems.

to change the I_{DS} by one decade. This quantity is important because it is needed in the design of digital circuits and because it is a sensitive indicator of the presence or absence of short-channel effects. Typical S factors in modern processes are around 80 mV/decade, so the maximum acceptable change, due to short-channel and high V_{DS} effects would be about 5 mV/decade; i.e., if the long-channel device exhibits $S = 80$ mV/decade, then the shortest-channel devices should have $S \leq 85$ mV/decade.

A good example of device characteristics with good electrostatic integrity are shown in Fig. 6.11b. As can be seen, the shortest-channel device exhibits a small change of S relative to the long-channel device. Also note that the short-channel device exhibits a shift to the left of the $\log I_{DS}$ - V_{GS} characteristic when V_{DS} is increased. This shift is due, as we have seen, to drain-induced barrier lowering (DIBL), which is typically characterized by $\Delta V_{GS}(\text{mV})/\Delta V_{DS}$ at constant I_{DS} in weak inversion. Typical values of DIBL for devices with good electrostatic integrity are less than 100 mV/V.

The characteristics of the two shorter devices in Fig. 6.11a fail the criteria of good electrostatic integrity because they exhibit strong V_{DS} dependence of their S factor. As already mentioned in Sec. 6.3, these devices suffer from the “punchthrough” effect.^{27,35,39,59,66–69,86,107–111} Punchthrough is a severe case of barrier lowering causing electron flow from source to drain either along the surface, in which case it is referred to as *surface punchthrough*, or in the bulk, referred to as *bulk punchthrough*. An example of what the I_{DS} - V_{DS} characteristics of a punchthrough-suffering device look like is shown in Fig. 6.19. These effects are easily modeled by 2-D device simulation but are not worth including in compact models, as they are not encountered in practical devices.

In a rather simplified way, surface punchthrough is said to occur when the depletion regions of the source and drain, in the absence of any depletion from the gate, reach each other. Figure 6.20 shows two such cases: (1) surface-punchthrough for the case of a uniformly doped substrate, as was the case for the devices in Fig. 6.11a, and (2) bulk punchthrough for the case of ion-implanted channel with higher concentration at the surface. It is worth pointing out that the signature of bulk punchthrough is primarily a “bottoming out” of I_{DS} with decreasing V_{GS} as shown in Fig. 6.21. Little change of S with V_{DS} is observed, but the I_{DS} “floor” is strongly V_{DS} dependent.

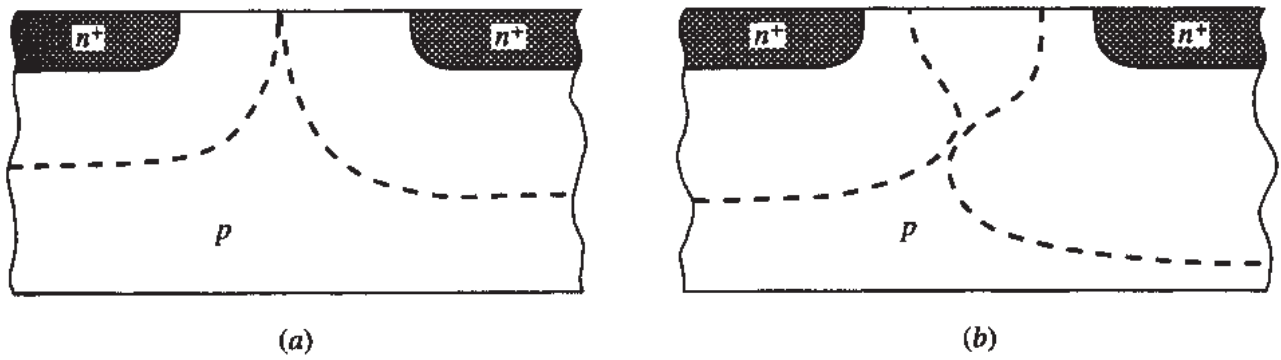
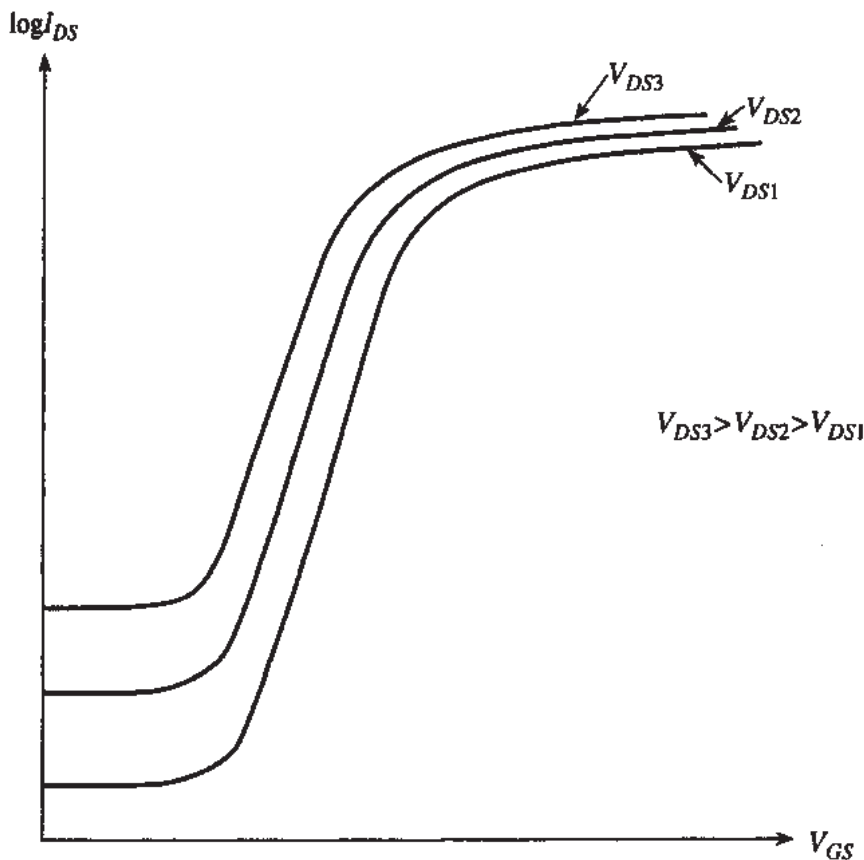
**FIGURE 6.20**

Illustration of source and drain depletion edges at the onset of punchthrough. (a) Surface punchthrough; (b) bulk punchthrough.

**FIGURE 6.21**

Log I_{DS} vs. V_{GS} for three different values of V_{DS} for a device exhibiting bulk punchthrough.

Although the notion of touching depletion regions is simplistic (for example, it ignores the fact that after all a depletion region always exists along the channel due to the gate), it does convey the idea that with a diminished neutral region between source and drain, field lines from the latter can reach the vicinity of the first and modulate the surface potential, and hence Q'_f . Clearly, as pointed out in Sec. 6.3, the remedy for punchthrough is to shrink the depletion regions by adding doping, usually through ion implantation. For example, in the case depicted in Fig. 6.20b a “punchthrough-control” implant can be introduced, in addition to the surface V_T -control implant. Its effect can be adjusted to just eliminate the I_{DS} bottoming out of Fig. 6.21.

6.5 CARRIER VELOCITY SATURATION

Up to this point in this chapter we have considered effects that are due to the electrostatic coupling of the source, and even more significantly of the drain, to channel charge, as the channel length decreases. These effects are manifest both in weak and strong inversion via a shift of \hat{V}_T . The fact that the gate in those devices does not have full control of the channel charge is a clear indication that the *longitudinal* electric field (i.e., parallel to the current flow direction; also called *parallel* or *tangential*) is not negligible compared to the *transverse* field (also called *normal* field) caused by the gate over a large portion of the channel. This means that effects associated with high longitudinal fields must be incorporated in our analysis. In terms of the drain current under normal device operation the most significant effect that needs to be included is an effective reduction of mobility with increasing longitudinal field, as we will see shortly. All the nonsaturation region models we have considered so far are based on an assumption stated above (4.3.2) concerning the electric field in the inversion layer. To recall that assumption, let \mathcal{E}_x be the value of the *longitudinal* field component. We have assumed that at all points in the inversion layer $|\mathcal{E}_x|$ is small enough so that the magnitude of the carrier velocity $|v_d|$ is proportional to $|\mathcal{E}_x|$. In devices with small-channel lengths this assumption is not accurate, and the I_{DS} - V_{DS} relations we have derived will not be valid. In this section we will show how adequate modeling can be achieved in such cases.^{18,23,24,27,34,36-40,112-118}

Figure 6.22 illustrates the behavior of $|v_d|$ with $|\mathcal{E}_x|$. As was the case for bulk conduction (Sec. 1.3), the velocity of carriers in the inversion layer tends to saturate at high $|\mathcal{E}_x|$ values.[†] Effects due to the lack of proportionality between $|v_d|$ and $|\mathcal{E}_x|$ on device characteristics are often referred to as *velocity saturation effects*, although the $|\mathcal{E}_x|$ values involved may be below those corresponding to the clear saturation part of the figure.

It is convenient to define a “critical” value of $|\mathcal{E}_x|$, denoted by \mathcal{E}_c , at the intersection of the $|v_d| = \mu|\mathcal{E}_x|$ line and an imaginary horizontal asymptote, as shown in the figure. We have

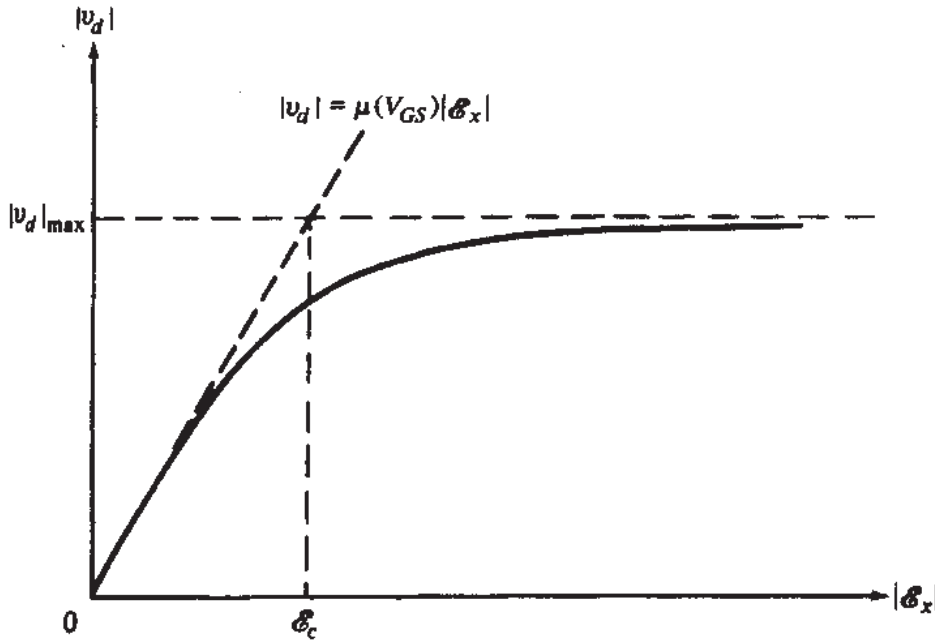
$$|v_d| \approx \mu|\mathcal{E}_x|, \quad |\mathcal{E}_x| \ll \mathcal{E}_c \quad (6.5.1a)$$

$$\approx |v_d|_{\max}, \quad |\mathcal{E}_x| \gg \mathcal{E}_c \quad (6.5.1b)$$

and

$$\boxed{\mathcal{E}_c = \frac{|v_d|_{\max}}{\mu}} \quad (6.5.2)$$

[†]The velocity can attain values larger than $|v_d|_{\max}$ over very short times or very short distances. This effect, referred to as *velocity overshoot*¹¹⁹⁻¹²¹, is not considered here. The physics behind velocity saturation are complicated; readers with adequate background in solid-state physics can consult Ref. 116. In general, the physics of carrier transport over very short distances is involved; even the expression of current as a drift plus a diffusion component has been questioned. For a summary of such considerations, the reader is referred elsewhere.³⁹ These considerations are in the realm of physics for numerical simulation and so far have not led to simple analytic device models.

**FIGURE 6.22**

Magnitude of carrier velocity in the inversion layer vs. magnitude of longitudinal component of electric field, $|E_x|$. $\mu(V_{GS})$ is the low $|E_x|$ field surface mobility at a given V_{GS} (see Sec. 4.10).

Values used for the parameters in the above equation vary. The values for $|v_d|_{\max}$ and E_c must, of course, be consistent with the value for μ , which is a function of the average transverse field $E_{y,ave}$, as we have seen in Sec. 4.10. "Transverse" indicates the direction perpendicular to the direction of current flow. In some treatments, the same value of $|v_d|_{\max}$ is used for both electrons and holes (5×10^6 to 2×10^7 cm/s),[†] which results in E_c values of about 8×10^3 to 3×10^4 V/cm for electrons and 2×10^4 to 10^5 V/cm for holes, assuming values of low transverse-field mobility of 650 and 230 $\text{cm}^2/(\text{V} \cdot \text{s})$, respectively. In other treatments $|v_d|_{\max}$ is taken in the above range for electrons but somewhat lower for holes (by a factor usually close to 1, but sometimes as large as 3), in which case E_c for holes is reduced accordingly. The reason for these discrepancies is that these parameters are difficult to measure, and that several different theories have been developed to explain them. Also, in MOS transistor work these theories are usually applied in a simplified manner. Hence, some empirical adjustment of numerical values is performed to improve overall accuracy in the resulting device models.

Various empirical relations have been used to model the dependence of $|v_d|$ on $|E_x|$. Thus, in some treatments the curve in Fig. 6.22 is replaced by a piecewise-linear plot, as suggested by the two asymptotes.²³ Another relation in use is^{112,114}

$$|v_d| = |v_d|_{\max} \frac{|E_x|/E_c}{1 + |E_x|/E_c} \quad (6.5.3)$$

[†]Experimentally, the value of $|v_d|_{\max}$ for electrons is found¹²² to be about 9×10^6 cm/s, while for holes about 8×10^6 cm/s, at room temperature. The broader range of values quoted above is necessary to compensate for model inadequacies.

This tends to model better the region between very low and very high fields and is still in agreement with (6.5.1). Although expressions more complicated than (6.5.3) can provide more accuracy,[†] this equation is widely used because it leads to a simple transistor model as will be seen.

Consider now operation in the nonsaturation region of the $I_{DS}-V_{DS}$ strong-inversion characteristics and assume, as in Sec. 4.5, that all current is due to drift. Let $V_{CB}(x)$ be the "effective reverse bias" of the inversion layer at point x with respect to the bulk, as in Sec. 4.5. The voltage drop across a piece of the inversion layer (say, of length Δx) is equal to the difference ΔV_{CB} between the V_{CB} values at the two ends of the piece. Letting the above finite differences become differentials, we have, for the magnitude of the field intensity at point x ,

$$|\mathcal{E}_x(x)| = \frac{dV_{CB}}{dx} \quad (6.5.4)$$

Thus, the magnitude of the electron velocity at point x becomes, from (6.5.3),

$$|v_d(x)| = |v_d|_{\max} \frac{(1/\mathcal{E}_c)(dV_{CB}/dx)}{1 + (1/\mathcal{E}_c)(dV_{CB}/dx)} \quad (6.5.5a)$$

or, from (6.5.2):

$$|v_d(x)| = \mu \frac{dV_{CB}/dx}{1 + (1/\mathcal{E}_c)(dV_{CB}/dx)} \quad (6.5.5b)$$

To calculate the drain current in nonsaturation I_{DSN} , we cannot use (4.5.6) anymore. That equation was based on (1.3.15), which was derived by assuming $|v_d|$ was proportional to $|\mathcal{E}_x|$. Since this is not the case anymore, we will use the more general expression (1.3.7). For the MOS transistor, Q' in that expression is the inversion layer charge per unit area Q'_I , the value of which depends on $V_{CB}(x)$ and will be denoted by $Q'_I[V_{CB}(x)]$, and b is the channel width W . Thus

$$I_{DSN} = W(-Q'_I) |v_d(x)| \quad (6.5.6)$$

Using (6.5.5b) in (6.5.6) we obtain

$$I_{DSN} \left(1 + \frac{1}{\mathcal{E}_c} \frac{dV_{CB}}{dx} \right) = \mu W(-Q'_I) \frac{dV_{CB}}{dx} \quad (6.5.7)$$

Integrating from $x = 0$ (where $V_{CB} = V_{SB}$) to $x = L$ (where $V_{CB} = V_{DB}$), and assuming μ and \mathcal{E}_c are independent of the transverse field as in Sec. 4.5, we obtain

[†]It has been suggested¹¹⁶ that, while (6.5.3) is adequate for holes, $|v_d|$ for electrons can more closely be approximated by $|v_d|_{\max} (|\mathcal{E}_x|/\mathcal{E}_c) / \sqrt{1 + (|\mathcal{E}_x|/\mathcal{E}_c)^2}$. Another model¹¹⁸ uses (6.5.3) for $|\mathcal{E}_x| \leq \mathcal{E}_c$, and $|v_d| = \mu \mathcal{E}_c/2$ for $|\mathcal{E}_x| > \mathcal{E}_c$; in this case, the value of \mathcal{E}_c to be used for best fit to experiment will be different from that used in conjunction with (6.5.3).

$$I_{DSN} \left(L + \frac{V_{DB} - V_{SB}}{\mathcal{E}_c} \right) = \mu W \int_{V_{SB}}^{V_{DB}} (-Q'_I) dV_{CB} \quad (6.5.8)$$

Using $V_{DB} - V_{SB} = V_{DS}$, we have

$$I_{DSN} = \frac{W}{L} \frac{\mu}{1 + V_{DS}/(L\mathcal{E}_c)} \int_{V_{SB}}^{V_{DB}} (-Q'_I) dV_{CB} \quad (6.5.9)$$

Comparing now this equation to (4.5.7), with μ assumed constant, we see that they are identical except for the multiplicative factor $1/[1 + V_{DS}/(L\mathcal{E}_c)]$ *outside* the integral. Consequently, nonsaturation region expressions developed in Chap. 4 can be made to be valid in the presence of velocity saturation effects by multiplying them with this factor. Thus

$$I_{DSN, \text{ including velocity saturation}} = \frac{I_{DSN, \text{ not counting velocity saturation}}}{1 + V_{DS}/(L\mathcal{E}_c)} \quad (6.5.10)$$

In the right-hand side of this equation one can use the accurate model of (4.5.2), the approximate model of (4.5.37), or any other model at hand, depending on the accuracy desired.

Example 6.2. Using the approximate model of (4.5.37a) in (6.5.10) we have for the current in the presence of velocity saturation effects,

$$I_{DS} = \frac{W}{L} \frac{\mu C'_{ox} [(V_{GS} - V_T)V_{DS} - 0.5\alpha V_{DS}^2]}{1 + V_{DS}/(L\mathcal{E}_c)}, \quad V_{DS} \leq V'_{DS} \quad (6.5.11)$$

The effect of normal field (Sec. 4.10) can also be incorporated in the above formulation. Assuming independence between mobility and saturation velocity,¹¹⁶ one can replace μ in (6.5.1) and (6.5.2) by (4.10.13), and rederive I_{DSN} using the approximations suggested in Sec. 4.10. One then obtains again the above expressions, only with μ replaced by the effective mobility μ_{eff} given by (4.10.16) or (4.10.20). Other approaches²⁷ give similar results.

The I_{DS} expressions developed as shown above can easily be seen to attain zero slope dI_{DS}/dV_{DS} at a V_{DS} value *smaller* than that found in the absence of velocity saturation effects. This suggests the following simplifying assumption, the validity of which will be considered shortly: the “saturation” of I_{DS} - V_{DS} curves will be assumed to be attained purely owing to velocity saturation effects. Thus, the value of V'_{DS} at which saturation occurs can be found by solving $dI_{DS}/dV_{DS} = 0$ as before.[†]

[†]Alternatively, one can equate (6.5.11) evaluated at $V_{DS} = V'_{DS}$ to (6.5.6) evaluated at the drain with $|v_d(L)| = |v_d|_{\text{max}}$, and solve for V'_{DS} . The result will be the same. A more exact calculation³⁷ aided by numerical simulation results, leads to a complicated expression for V'_{DS} which, however, results in practically the same value as the approach suggested here.

Example 6.3. For the model of Example 6.2 we have, setting dI_{DS}/dV_{DS} equal to zero and solving (Prob. 6.20),

$$V'_{DS} = \frac{V_{GS} - V_T}{\alpha} \frac{2}{\sqrt{1 + \left(\frac{V_{GS} - V_T}{\alpha}\right) \cdot \frac{2}{L\mathcal{E}_c}} + 1} \quad (6.5.12)$$

This value is smaller than the value of $(V_{GS} - V_T)/\alpha$ expected in the absence of velocity saturation effects. It only approaches that value as $L\mathcal{E}_c$ approaches infinity, i.e., in the absence of velocity saturation effects in the limit (Prob. 6.20).

In the saturation region ($V_{DS} > V'_{DS}$), the inclusion of channel length modulation effects is necessary. Continuing our simplifying assumptions from above, we postulate that in the channel region adjacent to the drain, where $V_{DS} - V'_{DS}$ is dropped, the electron density is small. Then the width l_p of this region can be found as in Sec. 6.2. Thus the drain current in saturation can be found from (6.2.5) and (6.2.6), as before, and I'_{DS} in that formulation is the value of I_{DS} at $V_{DS} = V'_{DS}$, where V'_{DS} is found as explained above. The value of V'_{DS} can be replaced by a slightly lower value \hat{V}_{DS} to allow continuity of the slope of dI_{DS}/dV_{DS} , as has been already discussed in conjunction with Fig. 6.3c.

Example 6.4. Using the appropriate model (4.5.37a) in the right-hand side of (6.5.10) with $V_{DS} = V'_{DS}$, where V'_{DS} is given by (6.5.12), and replacing L by $L - l_p$ as in Sec. 6.2, we have for the current in saturation, in the presence of velocity saturation effects,

$$I_{DS} = \frac{W\mu C'_{ox}[(V_{GS} - V_T)V'_{DS} - 0.5\alpha V'^2_{DS}]}{L\left(1 - \frac{l_p}{L} + \frac{V'_{DS}}{L\mathcal{E}_c}\right)} \quad (6.5.13)$$

In Ref. 37, the saturation region is at first investigated without making the simplifying assumptions above. Simplifications are then sought, using two-dimensional computer simulations as a guide. The final result consists of the same form of equations as the ones we found above, and these are shown to be in satisfactory agreement with experimental results.

Velocity saturation effects can have a drastic influence on I_{DS} - V_{DS} characteristics. Figure 6.23 compares results calculated, assuming such effects are absent (Fig. 6.23a) and present (Fig. 6.23b). It is seen that, for the same V_{GS} , saturation is achieved at smaller values of V_{DS} when velocity saturation is included. Most importantly, in that case the spacing of the curves in the saturation region is *not* according to the "square law" of (4.5.37b). In fact, if the velocity saturation effects are severe, the spacing becomes nearly proportional to the V_{GS} increment, as is the case in Fig. 6.23b.† We now have a device the saturation current of which depends almost lin-

† Actually, the mobility dependence on the vertical field, discussed in Sec. 4.10, also has the effect of "linearizing" the square-law dependence of I'_{DS} on V_{GS} . For modern MOSFETs with $t_{ox} < 100 \text{ \AA}$ and $L < 0.5 \text{ }\mu\text{m}$, I'_{DS} is almost linear with V_{GS} .

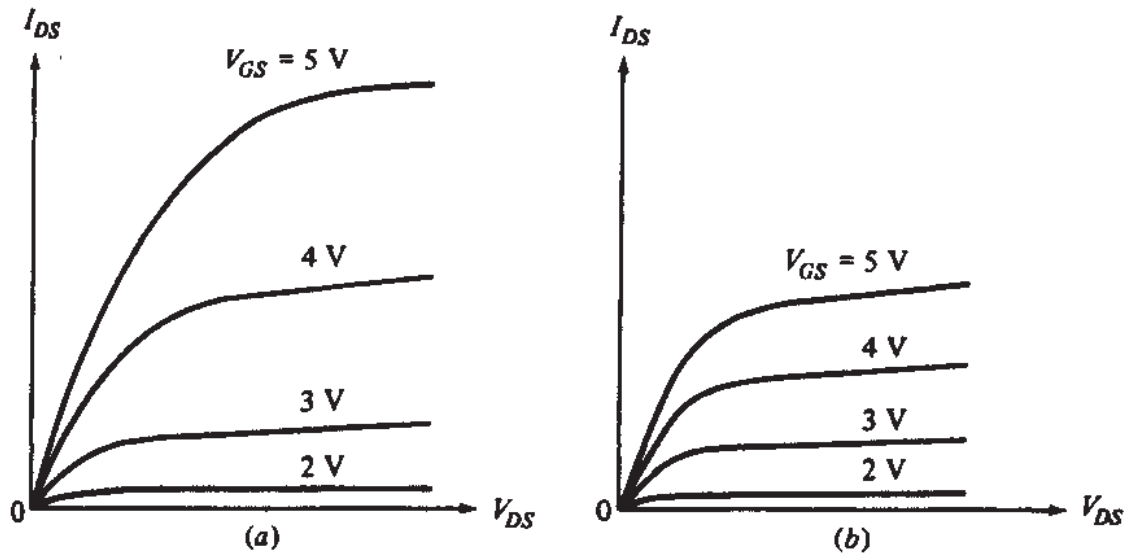


FIGURE 6.23

Device characteristics (a) in the absence and (b) in the presence of velocity saturation effects.

early on $V_{GS} - V_T$! This can be seen from the equations in Examples 6.3 and 6.4, as follows. The effects of velocity saturation become stronger as L is decreased. V'_{DS} decreases with decreasing L , although not as fast as L itself. When L is very small [small V'_{DS} from (6.5.12)], we can neglect the square term in the numerator of (6.5.13). Assuming for simplicity that l_p/L in the denominator is much less than 1, we have

$$I'_{DS} \approx \frac{\mu C'_{ox}(W/L)(V_{GS} - V_T)V'_{DS}}{V'_{DS}/(L\mathcal{E}_c)}, \quad \text{very small } L \quad (6.5.14a)$$

$$\approx WC'_{ox}(V_{GS} - V_T)\mu\mathcal{E}_c, \quad \text{very small } L \quad (6.5.14b)$$

The linear dependence of I'_{DS} on $V_{GS} - V_T$ is apparent. We can put this relation in an alternative form by postulating that, since V'_{DS} is small, the channel charge will be approximately uniform and independent of x , so that we have $-Q'_l \approx C'_{ox}(V_{GS} - V_T)$ at any point x . Using this fact and using (6.5.2) in (6.5.14b), we have

$$I'_{DS} \approx W(-Q'_l)|v_d|_{\max} \quad (6.5.15)$$

which could have been obtained directly from (6.5.6), assuming the carriers are traveling at maximum velocity from the moment they are injected by the source to the moment they are collected by the drain.

The above two relations also reveal another major effect in the limit of very strong velocity saturation: *The drain current is independent of L !* A physical feeling for this effect can be obtained as follows. The transit time of carriers in the channel (the time it takes them to travel the length of the channel) is proportional to L , assuming for simplicity that they travel at constant maximum velocity. The total mobile charge in the channel is also proportional to L . Hence the current, which is the ratio of this charge to the transit time, is independent of L . To put it another way, if the distribution of mobile charge is uniform and that charge moves at constant velocity, a

fixed amount of charge passes per unit time for a given W . Thus the drain current is fixed. The length of the channel does not enter in this reasoning at all (provided, of course, that L is small enough so that the velocity saturation effect assumed is valid in the first place). An analogy can be given in terms of water coming out of the end of a pipe at constant speed. The water flow rate out of the pipe is independent of the length of the pipe.

The approximations used in the preceding three paragraphs are actually oversimplified since, when the channel length is small, the two-dimensional field near the source and drain regions can have a significant influence on the channel charge, as we have already seen. The above discussion is, nevertheless, useful for providing intuition and for relating several of the results we have presented. Note that, because $|v_p|_{\max}$ has similar values for electrons and holes,^{115,122} n -channel and p -channel devices tend to perform similarly under velocity saturation, other things being equal. This is not the case in the absence of velocity saturation, e.g., for very long channels, since then I_{DS} is proportional to μ , the value of which for p -channel devices is one-third to one-fourth the corresponding value for n -channel devices. Indeed, it is generally observed in modern technologies that the ratio of n -channel current to p -channel current, for equal L and the same bias, is reduced as L is reduced and the effect of velocity saturation becomes more prevalent.

6.6 HOT CARRIER EFFECTS— SUBSTRATE CURRENT, GATE CURRENT, AND BREAKDOWN

As we saw in the previous sections, the longitudinal electric field in the channel increases from source to drain. For abrupt source and drain junctions, the peak field is at the drain-to-channel junction, and its value depends on V_{DS} and L . Figure 6.24 illustrates this for two devices of different L , at the same value of V_{DS} (an abrupt junction is assumed; see Appendix C). The peak value of the longitudinal electric field, \mathcal{E}_m , depends strongly on these parameters. Also shown in the figure is the approximate level \mathcal{E}_c above which the carrier velocity is expected to saturate. For the longer-channel device this almost coincides with the beginning of the pinchoff region, while for the shorter-channel device there is a portion of the channel before pinchoff where the carrier velocity saturates. When carriers move in fields that exceed the value of the onset of velocity saturation, they continue to acquire kinetic energy from the field but their velocity is randomized by excessive collisions such that their velocity along the field direction no longer increases but their *random* kinetic energy does. Depending on the statistics of scattering, a small fraction of the overall carrier population acquires a significant amount of energy, and these are called *hot carriers*. Clearly, the higher the field, the higher the proportion of hot carriers in the overall population. Generally, in MOSFETs, the high fields are encountered in saturation in the pinchoff region, as is evident from Fig. 6.24.

Figure 6.25 illustrates the *hot carrier effects*. “Cool” electrons are coming into the pinchoff region and are *heated* by the field. Some of them acquire enough energy to create *impact ionization* of silicon lattice atoms, whereby new electrons and holes are created; this effect is also referred to as *weak avalanche*. The new electrons join

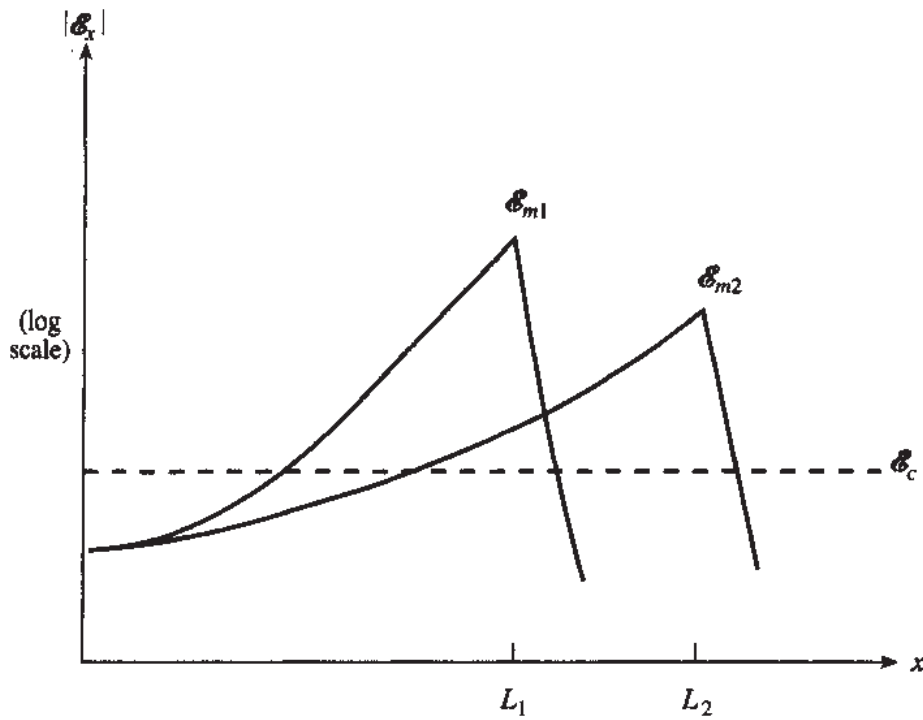
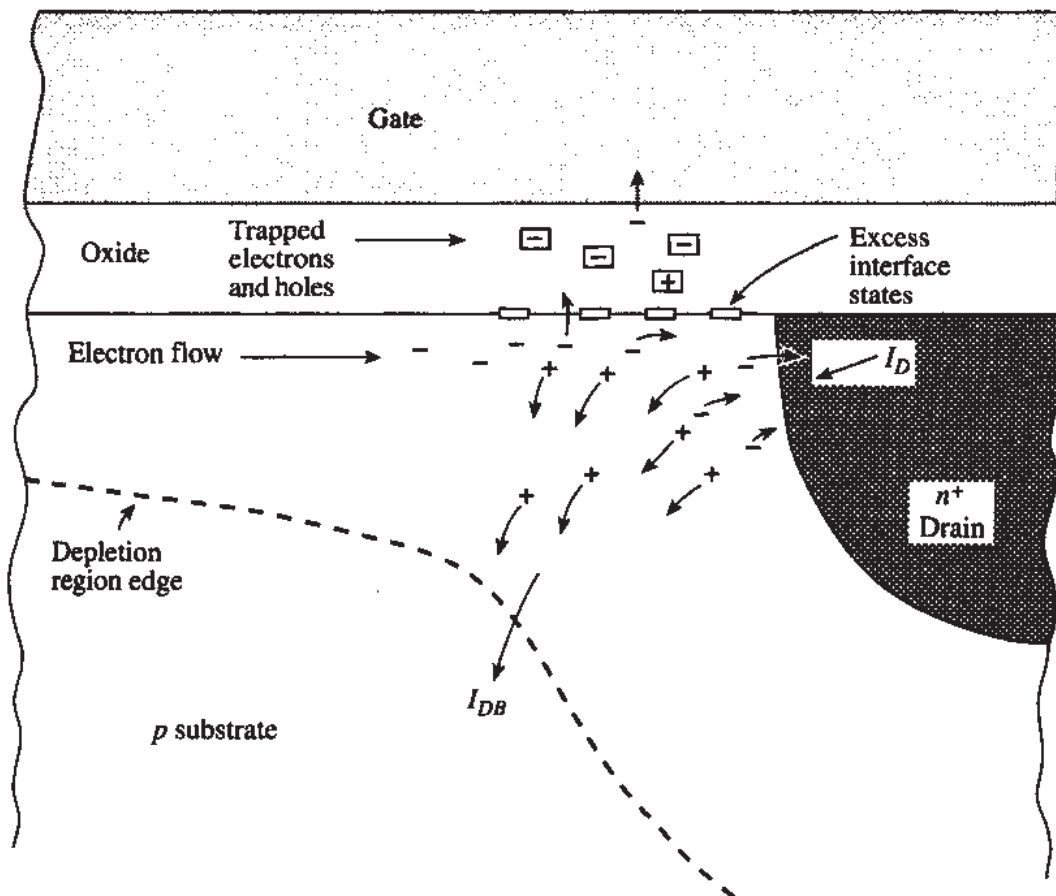
**FIGURE 6.24**

Illustration of $|E_x|$ (log scale) vs. x along the silicon to oxide interface for two transistors in saturation with the same V_{DS} . Note the smaller maximum electric field for the device with the longer channel length (L_2). For illustration purposes the field magnitude, E_c , at which velocity saturation occurs is also shown.

**FIGURE 6.25**

Schematic representation of hot carrier effects in a region of high longitudinal electric field in the channel of an nMOSFET.

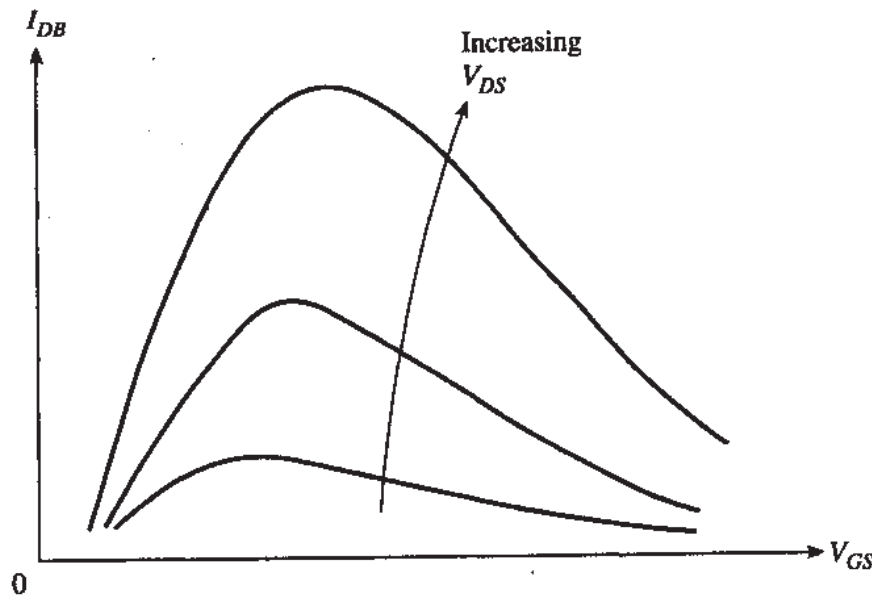


FIGURE 6.26

Substrate current vs. gate-source voltage, with drain-source voltage as a parameter.

the stream of channel electrons and move on toward the drain. The normal depletion field in the channel pushes the holes into the substrate, where they give rise to *drain-to-substrate current*, I_{DB} .¹²³⁻¹⁵⁸ A much smaller fraction of electrons acquire even higher energy which allows them to overcome the silicon-oxide barrier, get injected into the oxide, and be collected by the gate as *gate current*. Of those very energetic carriers, a small fraction create damage at the silicon-oxide interface which manifests itself as an increase in interface states density N_{it} (Sec. 2.2), and yet another small fraction become trapped in the oxide giving rise to a localized change of Q_o' .^{135,136,138,140-143} (see also Sec. 2.2). The ensuing corruption of the oxide results in device degradation with operating time, or "aging."

Let us consider the drain-to-substrate current I_{DB} . This current is proportional to the number of electrons available per unit time, which in turn is proportional to I_{DS} . Also, according to the above discussion, I_{DB} is an increasing function of the maximum field at the drain; this field is, in turn, a function of the "excess" drain voltage, $V_{DS} - V'_{DS}$.

Plots of I_{DB} vs. V_{GS} , for various values of V_{DS} , behave as shown in Fig. 6.26. For a given V_{DS} , when V_{GS} is increased starting from low values, I_{DS} increases, and thus I_{DB} increases too, according to the above discussion. Further increases of V_{GS} increase V'_{DS} significantly, causing a strong decrease in $V_{DS} - V'_{DS}$, and thus in the maximum field at the drain. This causes I_{DB} to decrease at large V_{GS} values, as seen toward the right in the figure. The maximum I_{DB} is observed, roughly, at $V_{GS} \approx V_{DS}/2$.

Developing an analytical expression for I_{DB} is a lengthy process; after several approximations, it can be shown that I_{DB} is of the form,^{88,140,158}

$$|I_{DB}| = |I_{DS}| K_i (V_{DS} - V'_{DS}) \exp \left(-\frac{V_i}{V_{DS} - V'_{DS}} \right) \quad (6.6.1)$$

where K_i and V_i are parameters that must be fitted empirically to data. This form has been corroborated by experimental data¹⁵⁸ and is well accepted at present. Typical values for K_i are 1 to 3, and for V_i , 10 to 30 V. In other models¹⁴⁸ for the substrate current, the $(V_{DS} - V'_{DS})$ term is replaced by a constant, and the factor 0.8 is introduced in front of V'_{DS} in the exponent. The *total* drain current, I_D , is given by

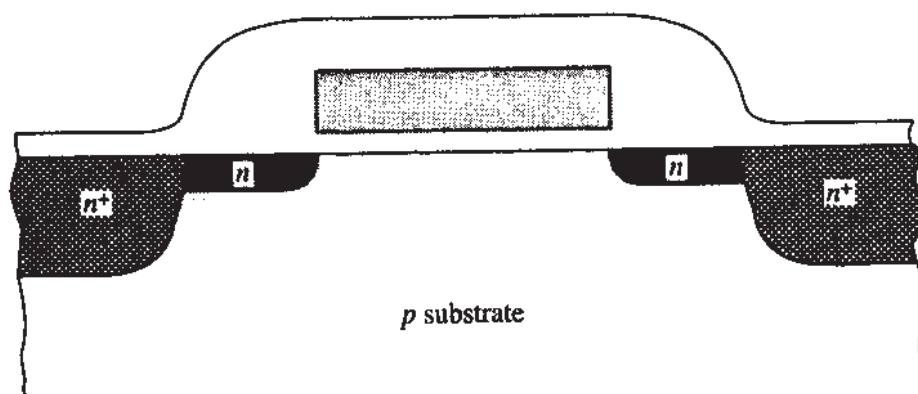
$$I_D = I_{DS} + I_{DB} \quad (6.6.2)$$

The discussion so far about substrate current has been based on the assumption that the latter is small compared to drain current, for example, more than 2 to 3 orders of magnitude smaller. However, as V_{DS} increases to well above V'_{DS} , depending on the device design, it reaches a point at which the substrate current becomes comparable to drain current. At that point the *total* drain current I_D is seen to increase rapidly with V_{DS} . This condition marks the onset of considerable avalanche multiplication, in which the newly created carriers can participate in significant numbers in generating more pairs as they gather energy from the field; the device is then said to be operating in the breakdown region. In addition, a significant hole current can cause enough ohmic voltage drop across the resistance of the substrate material to forward-bias the substrate-source *pn* junction. For sufficiently high forward-bias, electrons can then be injected by the source into the substrate, and a parasitic bipolar transistor action can be initiated, contributing to breakdown. Note that breakdown in MOSFETs, as in diodes, is a reversible effect; i.e., the device is not necessarily destroyed. However, at this condition of excessive V_{DS} MOSFETs degrade (age) rapidly because of the large population of hot carriers impinging on the silicon-oxide interface, as discussed above. Generally, MOSFETs are never operated in this region.

Models also exist in the literature for the gate current and device aging as a result of hot carriers.^{151,153,154,158,88} While substrate current can be significant enough to warrant modeling under normal device operation, the gate current is typically very small. As gate oxide thickness decreases, hot-carrier-generated gate current can become nonnegligible. Nevertheless, the limit of oxide thickness is considered to be set by what is called *direct tunneling*, as discussed in Sec. 6.9, rather than by hot carriers. One of the reasons for this is that, as oxide thickness is scaled down, so is the power supply voltage, and as a result hot carrier effects are rapidly diminished, since V_{DS} appears in an exponential [see (6.6.1)]. The same is true for the device aging rate, which is rapidly reduced as power supply voltage is scaled down.

In our discussions of various other phenomena in this book, the substrate and gate currents will be assumed negligible unless specifically stated otherwise.

To limit hot electron effects, it is advantageous to have a lightly doped drain region.¹⁵⁹⁻¹⁶⁴ This is because, in such a case, part of the depletion region would be inside the drain, absorbing some of the potential that otherwise would exist in the pinchoff region, and lowering the maximum electric field. However, a lightly doped drain in standard processes would mean a lightly doped source, since both regions are made simultaneously, in the same way. This would result in a large source resistance on which a large voltage drop could develop, which is undesirable, as will be discussed in Sec. 6.8. A compromise can be reached by using the structure shown in Fig. 6.27. This popular structure is referred to as *lightly doped drain* (LDD).^{159,161} It

**FIGURE 6.27**

Lightly doped drain (LDD) transistor; the n^+ to n doping ratio is between 10 and 100.

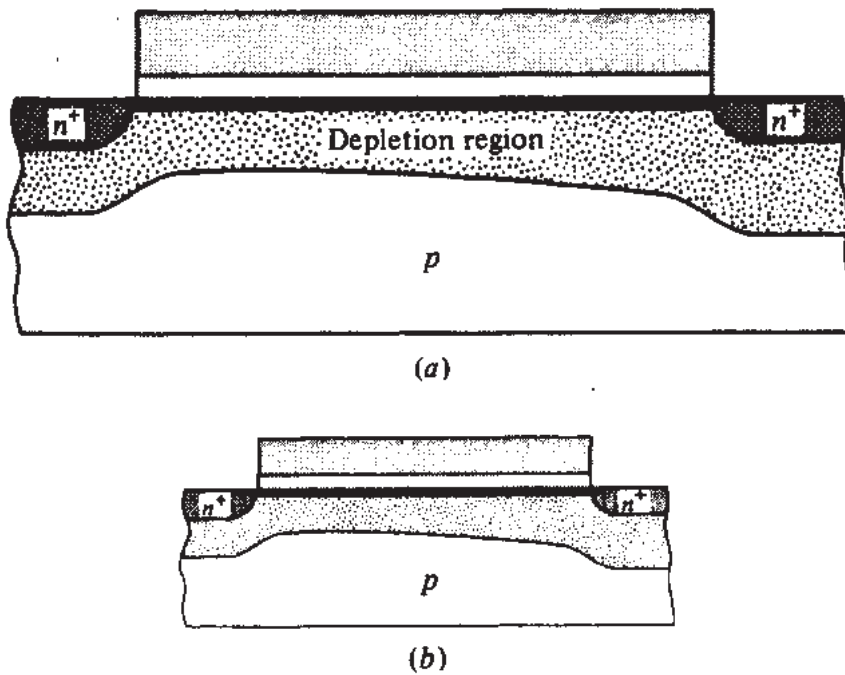
makes possible a lower maximum field, and thus less severe hot electron effects, while keeping the series resistance relatively low. In addition, because the lightly doped part can be made shallow, the short-channel effects related to charge sharing (Sec. 6.3.2) are limited.

6.7 SCALING

One of the hallmarks of the semiconductor industry is the continuous scaling of circuit dimensions with the introduction of every new technology generation.¹⁶⁵ This leads to a continuous increase in speed and circuit complexity per unit of chip area. However, if the channel length is made too small relative to the depletion regions around the source and drain, the short-channel effects associated with charge-sharing (Sec. 6.3) and punchthrough (Sec. 6.4) can become intolerable. Thus, to make L small, the depletion region widths should be made small. This can be done by increasing the substrate doping concentration and decreasing the reverse bias. To achieve the latter under any circuit operating conditions, the supply voltage must be decreased. Increasing the doping concentration increases the threshold voltage and makes it more difficult to turn the device on; this can be overcome by decreasing the oxide thickness.

As in the electrostatic integrity considerations above, maintaining the reliability of devices, i.e., constant aging rate, requires containment of the maximum electric field \mathcal{E}_m in the channel, discussed in Sec. 6.6. This is also achieved by reducing the maximum voltage, and by modifying the drain-channel junction to control \mathcal{E}_m .

What has just been described, i.e., adjusting a fabrication process and the bias voltages to allow proper operation of reduced-size devices, is one of the constant pursuits of fabrication process engineering.^{166–176} The adjustments aim at achieving small dimensions while, at the same time, avoiding severe side effects, such as the several small-dimension effects already considered in this chapter. Depending on what aspects of a given process should be optimized for an application in mind, there are many ways such adjustments can be made. Several of these are discussed in this section.

**FIGURE 6.28**

(a) A MOS transistor; (b) a MOS transistor obtained from (a) by scaling all geometrical dimensions by the same factor.

We will first describe a set of rules aimed at reducing size in such a way that the resulting device along with its depletion region is simply a scaled version of a large device, as is shown in Fig. 6.28. Then no significant side effects should appear, and one can analyze the scaled device by using well-known large-device concepts, thus taking advantage of the significant experience gained with large devices over the years. A process proposed to achieve the reduction shown in Fig. 6.28b will now be briefly described.¹⁶⁶ It will be seen to result in a “scaled” device, in which internal electric field shape and maximum magnitude is the same as in the original device; hence the name *constant-field scaling* is used to describe this process.

Let us assume that a large device is scaled in all three dimensions by a factor $1/\kappa$, where κ is larger than 1 (say, between 1 and 10). This means that L , W , oxide thickness, and junction depth are all scaled by the same factor. Thus, areas (e.g., gate area) are scaled by $(1/\kappa)^2$. The depletion region underneath a junction has a depth d given by [see (1.5.12) and (1.5.13)].

$$d = \sqrt{\frac{2\epsilon_s}{qN_A}} \sqrt{\phi_{bi} + V} \quad (6.7.1)$$

where V is the reverse bias and ϕ_{bi} the built-in potential. The form of this equation is also valid for the depletion region underneath the inversion layer; for example, in strong inversion the equation is valid with $\phi_{bi} = \phi_0$. We will assume ϕ_{bi} is small compared to V .† Then we can scale d by $1/\kappa$, by scaling $\sqrt{V/N_A}$ by $1/\kappa$ in the above

†This assumption is, of course, not valid for all bias conditions; it is made for simplicity. Thus, a device scaled as described will deviate from ideal constant field scaling behavior when the reverse bias is small.

equation. Let us scale N_A by κ and V by $1/\kappa$. To achieve a properly scaled V throughout the channel, all operating voltages as well as the threshold voltage must be scaled by $1/\kappa$. From basic electrostatics (Appendix B) it is easy to see that under this scaling the shape and maximum magnitude of the electric field in the structure will remain the same. Thus undesirable high field effects will not occur. As seen in (1.5.20) and (2.2.4) capacitances *per unit area* C' are inversely proportional to distances, so they are scaled by κ . However, capacitances C *per se* are given by $C'A$, where A is the area, and thus are seen to scale by $\kappa(1/\kappa^2) = 1/\kappa$.

The body effect coefficient γ is seen from (2.5.16) to scale by $1/\sqrt{\kappa}$. In device equations, the threshold voltage V_T appears in differences with bias voltages, such as in (4.5.3). Since the latter are scaled by $1/\kappa$, V_T should also be scaled by $1/\kappa$ as already mentioned. Consider (4.5.32). From the above discussion, if ϕ_0 is very small compared to V_{SB} , the term $\gamma\sqrt{\phi_0 + V_{SB}}$ will scale by $1/\kappa$. To scale $V_{FB} + \phi_0$ in that equation one should be able to control V_{FB} , which cannot be done independently for unimplanted devices. However, it has been seen in Chap. 5 that an "equivalent" V_{FB} can be controlled through ion implantation.

Charges *per unit area* Q' are not scaled [see (1.5.13) or (3.4.9) and use the above results]. Therefore, charges Q are scaled by $1/\kappa^2$, since areas are scaled by this factor.

What is now the effect of the above scaling procedure on the drain current? Let us look at (4.5.37). If α is close to 1, it is seen that since all voltages are scaled by $1/\kappa$, the quantity in brackets is scaled by $1/\kappa^2$. At the same time C'_{ox} is scaled by κ . Thus, recalling that μ is practically independent of doping concentration (Sec. 4.9) and thus does not change under constant-field scaling, the current in (4.5.37) is scaled by $1/\kappa$.

Consider now the slope of $\ln I_{DS}$ vs. V_{GS} in weak inversion for a constant V_{DS} . As seen from (4.6.13), that slope is proportional to $1/n$, with n given by (4.6.16). Since γ scales by $1/\sqrt{\kappa}$ and $V_{SB} + 2\phi_F$ by $1/\kappa$ (assuming V_{SB} is large), n and S in Sec. 4.6 remain the same, and the slope of $\ln I_{DS}$ vs. V_{GS} does not scale. For digital circuits, this is undesirable since it makes it more difficult to turn a device off. For example, it takes as much reduction in gate voltage to reduce I_{DS} by one decade as it does for large devices. Since the total voltage swings possible have been scaled by $1/\kappa$, the gate-voltage reduction needed to turn the device off represents a larger fraction of the total swing, and noise margins in digital circuits are reduced. The width of the moderate-inversion region¹⁷⁷ does not scale either. Thus this region now becomes a larger part of the power supply voltage, and more attention must be paid to it.^{177,178†}

Since both voltages and currents are scaled by $1/\kappa$, power dissipation is scaled by $1/\kappa^2$. However, since device areas have been scaled by $1/\kappa^2$, the density of devices per unit area is scaled by κ^2 ; thus, the power per unit of chip area is not scaled.

†In moderate inversion the slopes of the inversion layer charge (Fig. 2.10) and of I_{DS} for low V_{DS} (Fig. 4.28) vs. gate voltage are found to decrease. This has been attributed to the finite-inversion-layer capacitance in this region¹⁷⁷ [see (2.6.20)], and, for this reason, sometimes moderate-inversion effects are referred to as "finite-inversion-layer capacitance effects."

TABLE 6.1
Constant-field scaling

| Quantity | Scaling factor |
|---|-------------------|
| Device dimensions (L, W, t_{ox}, d_j) | $1/\kappa$ |
| Area | $1/\kappa^2$ |
| Packing density (devices per unit of chip area) | κ^2 |
| Doping concentration, N_A | κ |
| Bias voltages and V_T | $1/\kappa$ |
| Bias currents | $1/\kappa$ |
| Power dissipation for a given circuit | $1/\kappa^2$ |
| Power dissipation per unit of chip area | 1 |
| Capacitances, C | $1/\kappa$ |
| Capacitances per unit area, C' | κ |
| Charges, Q | $1/\kappa^2$ |
| Charges per unit area, Q' | 1 |
| Electric field intensity | 1 |
| Body effect coefficient, γ | $1/\sqrt{\kappa}$ |
| Transistor transit time, τ | $1/\kappa$ |
| Transistor power-delay product | $1/\kappa^3$ |

Since all device currents are scaled by $1/\kappa$ and the various capacitances are scaled by the same factor, the rate of change of charging these capacitances, $dV/dt = I/C$, will not scale. However, these capacitances now only need to be charged to voltages scaled down by $1/\kappa$, and thus the time needed to charge them scales by $1/\kappa$; hence, the speed of digital circuits increases by κ . Since the power dissipation for a transistor has scaled by $1/\kappa^2$, the “power delay product” (a figure of merit used for digital circuits) scales by $1/\kappa^3$. Constant-field scaling is summarized in Table 6.1.

Consider now the metal and polysilicon lines used to form the gates and interconnections. Since we now have a fabrication process capable of small dimensions, let us attempt to scale the width of these lines by $1/\kappa$. The new process may also require reducing the height of these lines, since trying to make very thin, but tall, lines can run into fabrication problems; let us attempt to scale that height by $1/\kappa$ also. Then the cross-sectional area of the lines is scaled by $1/\kappa^2$. Since the current that these lines carry has been seen to scale by $1/\kappa$, the current *density* in these lines will scale by κ . This is very undesirable since the increased current density can cause what is known as electromigration, a phenomenon in which atoms are carried by the flow of current and can result in line failure. For aluminum lines, the current density should not be larger than about $1 \text{ mA}/\mu\text{m}^2$. Another problem with scaling interconnection lines is that the resistance of the lines is proportional to length and inversely proportional to the cross-sectional area, and thus scales by κ . The parasitic capacitances of these lines to the substrate scale by $1/\kappa$, and thus the corresponding time constant does not scale. If the lines are long, this can cause a problem since it prevents us from taking advantage of the fact that the speed of transistors has scaled by κ . Also, since the resistance of these lines scales by κ and the current through them by $1/\kappa$, the voltage drop across them does not scale. Thus, a larger fraction of the total voltages available, which have been scaled by $1/\kappa$, is now wasted across interconnect lines. Because of the above problems, the height of interconnect lines is reduced less significantly.

Additional problems are caused by the "contact windows," etched through the thick oxide in order to make contacts between various layers. If the area of these windows is scaled by $1/\kappa^2$, their resistances will scale by κ^2 . For currents scaled by $1/\kappa$, this means that the voltage drop across the contacts will scale by κ , i.e., in the *opposite* direction from the bias voltages, which were scaled by $1/\kappa$. Another undesirable effect of scaling is an increase in the resistance of the source and drain n^+ regions due to the decreased junction depth. The effect of the source and drain resistances on the transistor characteristics is considered in Sec. 6.8.

The scaling scenario discussed above presents certain problems. As already discussed, the weak-inversion region width does not scale. Hence, the voltage swings required for turning the device from off to on may be an unacceptably large fraction of the total voltage available. In addition, established chip interface requirements must often be obeyed for which voltage levels are fixed, and, hence, the voltage cannot be scaled. The rules followed in such cases, for decreasing device dimensions while keeping the voltages unchanged, are referred to as *constant-voltage scaling*. In these, W , L , and N_A are scaled as before. However, if the oxide thickness is scaled by the same factor, the resulting field can be exceedingly high since voltages are not scaled; this can cause mobility degradation (Sec. 4.10). To alleviate this problem somewhat, oxide thickness is usually scaled less drastically.¹⁷⁶ One column in Table 6.2 summarizes these scaling rules. The resulting effect on the various quantities (such as those entered in Table 6.1) is considered in Prob. 6.23. It is, of course, to be expected that undesirable effects associated with high electric fields can become severe under constant-voltage scaling.

To avoid the extreme cases of constant-field and constant-voltage scaling, compromise scaling rules have been proposed. For example, geometric dimensions and substrate doping are scaled as in the case of constant-field scaling, but voltages are scaled less drastically. This has been termed *quasi-constant-voltage scaling*,¹⁸⁴ and is summarized in the corresponding column of Table 6.2. As is evident from (6.7.1), under this type of scaling the depletion region widths do not scale by the same factor as W , L , and t_{ox} . This is avoided if the scaling factor for N_A is modified appropriately, resulting in the *generalized scaling* rules shown in the last column of Table 6.2.¹⁷²

TABLE 6.2
Scaling rules

| Quantity | Scaling factor | | | |
|----------|------------------------|--|--|---|
| | Constant-field scaling | Constant voltage scaling $1 < \kappa' < \kappa$ | Quasi-constant voltage scaling $1 < \kappa' < \kappa$ | Generalized scaling $1 < \kappa' < \kappa$ |
| W, L | $1/\kappa$ | $1/\kappa$ | $1/\kappa$ | $1/\kappa$ |
| t_{ox} | $1/\kappa$ | $1/\kappa'$ | $1/\kappa$ | $1/\kappa$ |
| N_A | κ | κ | κ | κ^2/κ' |
| V, V_T | $1/\kappa$ | 1 | $1/\kappa'$ | $1/\kappa'$ |

A different approach concentrates on decreasing device dimensions while maintaining long-channel behavior in the weak-inversion region, such behavior being a sensitive indicator of the absence of severe short-channel effects. It has been found empirically that the minimum channel length for which such behavior is maintained fits the following relation:¹⁶⁹

$$L_{\min} = (\text{const}) [d_j t_{\text{ox}} (d_{BS} + d_{BD})^2]^{1/3} \quad (6.7.2)$$

where d_j is the source and drain junction depth, t_{ox} the oxide thickness, d_{BS} and d_{BD} the depletion region widths at the source and drain, respectively, and (const) has the value of $8.8 \mu\text{m}^{-1/3}$. The above relation portrays the various compromises involved and has proved to be a useful guide for decreasing device dimensions.

Before closing this section, let us speculate as to what might eventually be achieved^{170,171,173,179-189} by the ongoing effort to decrease physical dimensions and power supply voltages. We will try to present what might be considered an "average" view. However, predictions for the "future" concerning MOS devices have been repeatedly proved wrong in the past. Thus, one should not be surprised if the same were to happen with the predictions that follow.

The power supply voltage of an integrated circuit cannot be reduced arbitrarily because the correspondingly reduced signals become vulnerable to noise. For practical processes, also, the uncertainties of achieving a desired threshold voltage value are of the order of 0.1 V. In logic circuits the swings should be at least a few times that amount to ensure that all devices can be turned on and off with certainty. The last two considerations place a lower limit on the power supply voltage of 0.5 V or so, but speed considerations and driving requirements of off-chip components can make this value higher. This places a lower limit on device dimensions, to avoid breakdown. Other limitations can be imposed by the requirement to avoid overheating of the chip. Forced-air cooling allows about 20 to 40 W/cm² of chip area to be dissipated without overheating. Liquid cooling could increase this by an order of magnitude or even more by using special techniques.^{173,182} The degree of cooling imposes a limit on the number of transistors that can be placed in a given area if they dissipate significant power. For dynamic circuits, the need to avoid overheating has another repercussion. A circuit charging and discharging capacitor C to a voltage V , f times per second dissipates a power of fCV^2 . If many such circuits are closely packed, then, in order not to exceed the maximum power dissipation per unit of chip area, the maximum allowable clock frequency f may have to be limited.

Considerations such as the above can be found in several references dealing with fundamental limits.^{173,179,189} Taking the various factors into account, predictions as to what an "ultimate" MOS technology might look like vary. Devices with channel lengths as low as $0.04 \mu\text{m}$ have been demonstrated. Maximum "packing" densities predicted^{179,182} are in the range of 10^8 transistors per cm². Minimum switching times of inverters, "packed" as allowed by thermal limitations could be about 10 ps.¹⁸² Finally, because of delays encountered in propagating a signal over interconnection lines, clock frequency for a large synchronous digital network could be limited to about 3 GHz.¹⁸² Even more aggressive numbers have been predicted recently¹⁶⁵ using

extrapolations of historical trends of devices in production. Again, the above numbers have not been achieved in a production environment. They tend to indicate, though, what may be possible to achieve, and suggest that there is much room for improvement in VLSI.

6.8 EFFECT OF SOURCE AND DRAIN SERIES RESISTANCES

The MOS transistor channel is in series with two parasitic resistances, one associated with the source and the other with the drain. As shown in Fig. 6.29, one can identify three contributions to each such resistance: (1) The resistance R_1 of the metal contact to the n^+ region, (2) the resistance R_2 of the main body of the n^+ region (and n-LDD region if present, as in Fig. 6.27), and (3) the resistance R_3 associated with the crowding of the current flow lines, as they go from the n^+ region to the normally thinner inversion layer (the "spreading resistance"¹⁹⁰⁻¹⁹⁴). As new technology generations are developed, the junction depth is decreased, and all three components of the series resistance increase. One must then take the effect of the total resistance, $R = R_1 + R_2 + R_3$, into account. The reader is referred to the literature^{148,190-194} for the explanation of R which, especially in lightly doped drain (LDD) devices can exhibit significant voltage dependence.¹⁴⁸ In the following, we assume for simplicity that R is constant.

The effect of R on the drain current can be taken into account as shown in Fig. 6.30. The effective drain-source voltage \tilde{V}_{DS} is reduced below the voltage V_{DS} applied at the external terminals by the voltage drops across the series resistors:

$$\tilde{V}_{DS} = V_{DS} - 2RI_{DS} \quad (6.8.1)$$

We will use (4.5.37a) to obtain the drain current, with V_{DS} replaced by \tilde{V}_{DS} . For simplicity, we will assume that RI_{DS} is much smaller than $V_{GS} - V_T$. Then we do not need to consider the effective reduction in the gate-source voltage. Also, we will assume that \tilde{V}_{DS} is much smaller than $V_{GS} - V_T$, so that we can neglect the square term in (4.5.37a). Thus we have

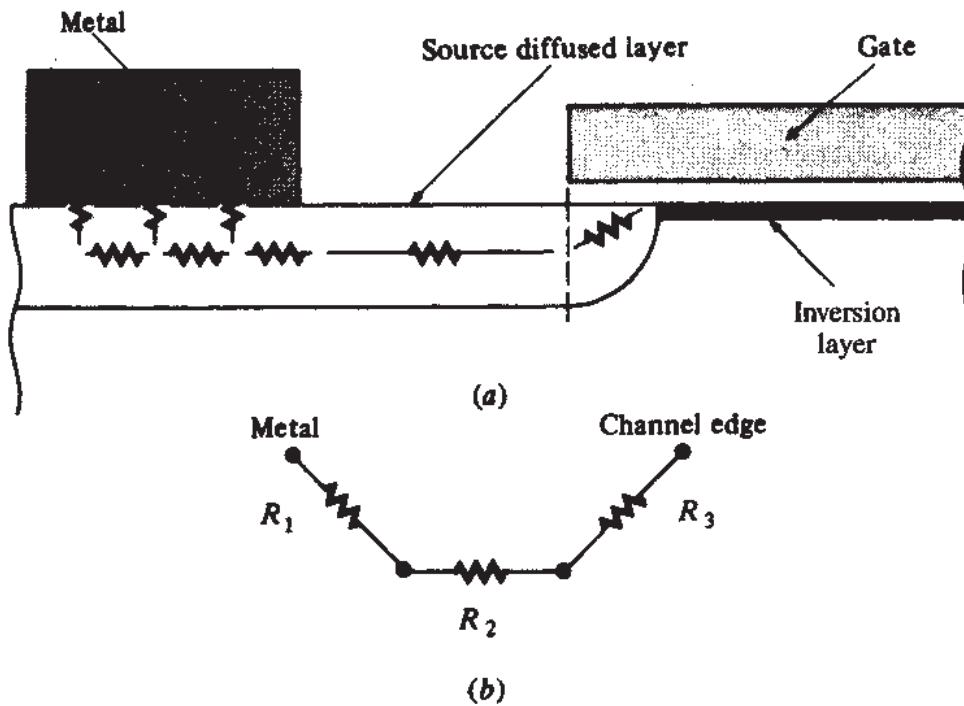
$$I_{DS} \approx \frac{W}{L} \mu C'_{ox} (V_{GS} - V_T) \tilde{V}_{DS} \quad (6.8.2)$$

Substituting (6.8.1) in this and solving for I_{DS} we obtain

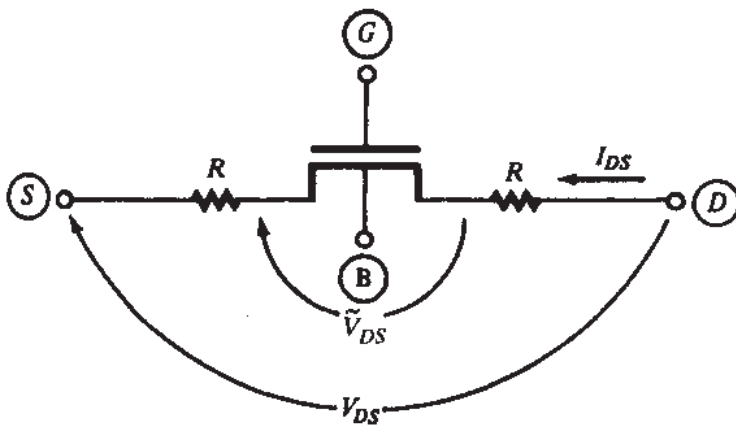
$$I_{DS} = \frac{\mu C'_{ox} (W/L)}{1 + \beta_R (V_{GS} - V_T)} (V_{GS} - V_T) V_{DS} \quad (6.8.3)$$

with

$$\beta_R = \frac{2\mu C'_{ox} RW}{L} \quad (6.8.4)$$

**FIGURE 6.29**

The components of the source series resistance in a MOSFET. (a) Structural correspondence; (b) lumped model.

**FIGURE 6.30**

A MOS transistor with the associated source and drain resistances.

In conservative fabrication processes, deep junctions, thick oxides, and large contact windows are used; then $C'_{ox}RW$ is small. In addition, L is large. Thus $\beta_R \approx 0$ can be used in (6.8.3), which implies that the current will be about the same as what would be observed if the series resistances were replaced by short circuits. In aggressive, short-channel fabrication processes, though, β_R is not negligible and the series resistance effect must be taken into account. The $I_{DS}-V_{GS}$ characteristic obtained from (6.8.3) is of the same form as that caused by the effective mobility reduction with V_{GS} , illustrated in Fig. 4.28. If we assume that both effects are present,

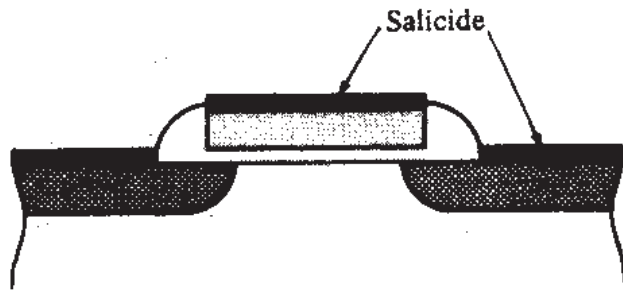


FIGURE 6.31
Transistor made by a salicide (self-aligned silicide) process.

then we should replace μ in (6.8.2) by an effective mobility expression from Sec. 4.10. Let us use (4.10.20) with $\theta_B = 0$ for simplicity. Then it is easy to show that, if $\theta(V_{GS} - V_T)$ and $\beta_R(V_{GS} - V_T)$ are both sufficiently smaller than 1, the factor $\mu/[1 + \beta_R(V_{GS} - V_T)]$ in (6.8.3) should be replaced by $\mu_0/[1 + (\theta + \beta_R)(V_{GS} - V_T)]$. This has led to the confusing usage of terms like “mobility reduction due to series resistance.” Such terms do not properly describe what happens. As is clear from our analysis, the two effects are totally separate. The fact that both happen to contribute to a term proportional to $V_{GS} - V_T$ in the denominator of the current expression is because of a mathematical coincidence.

A similar analysis can be carried out without making assumptions as to the relative magnitude of the various voltages involved. It is found, though, that even in such cases it is often sufficient to model the series resistance effect as above, i.e., by adding β_R to θ .

To keep the source/drain (and also the gate) resistances low, these regions are covered with metals (e.g., Ti, Ta, or Co) which are made to react with silicon and form a “disilicide.”¹⁹⁵ This process results in the device shown in Fig. 6.31 and is referred to as a *self-aligned silicide* (or *salicide*) process. It can reduce the sheet resistances by a factor or 5 to 10.

6.9 EFFECTS DUE TO THIN OXIDES AND HIGH DOPING

From the discussion in Secs. 6.3, 6.4, and 6.5, it is clear that in scaling down the lateral dimensions of MOSFETs, such as L and W , it is necessary to decrease the oxide thickness and increase the channel doping, so as to maintain the *electrostatic integrity* of the device. As an example, in technology generations circa 1998 with gate length of $0.25\ \mu\text{m}$ (i.e., electrical channel length L around $0.18\ \mu\text{m}$) oxide thickness ranges between 35 and $50\ \text{\AA}$, and channel doping is around $5 \times 10^{17}\ \text{cm}^{-3}$. Later generations have even thinner oxides and higher doping. We consider briefly in this section effects that become increasingly significant due to these scaling trends and limit the performance of devices. They are (1) decrease of the *effective oxide capacitance* due to the finite thickness of the inversion or accumulation layer and to depletion of the doped polysilicon gate; (2) increase of the threshold voltage due to *quantum mechanical* (QM) effects in the inversion (and accumulation) layers; and (3) *quantum mechanical* tunneling of carriers through very thin oxides.

INCREASE OF GATE OXIDE EFFECTIVE THICKNESS. So far we have always assumed that the gate controls the channel charge via a capacitance that is determined strictly by the physical thickness of the gate insulator, t_{ox} (assumed to be silicon dioxide, although it could be a different material such as silicon nitride or a combination of oxide and nitride). This is a consequence of the *charge sheet approximation* discussed in Sec. 4.3, where the inversion (or accumulation) layer is assumed to be infinitely thin. However, in reality the inversion charge is more spread out than an infinitely thin sheet at the surface. To properly calculate the shape of the inversion (and accumulation) region one has to solve simultaneously Poisson's equation, which governs electrostatics, and Schrödinger's equation, which governs the behavior of tightly confined particles.¹⁹⁶ When this is done, it is found that the distribution of carriers near the surface peaks at some depth d_m from the surface. Therefore the centroid of the inversion (or accumulation) charge distribution is away from the surface by a distance d_m . Then, the effective, or as it is often called, the *electrical*, oxide thickness which determines the capacitance coupling between the gate and the channel charge becomes

$$\hat{t}_{\text{ox}} = t_{\text{ox}} + \frac{\epsilon_{\text{ox}}}{\epsilon_s} d_m \quad (6.9.1)$$

where \hat{t}_{ox} is now the electrical oxide thickness that should be used in all model equations in place of t_{ox} . Note that the change of dielectric coefficient at the silicon oxide interface is accounted for by the ratio $\epsilon_{\text{ox}}/\epsilon_s$. The depth d_m depends on the confining transverse field near the surface, $\mathcal{E} = d\psi/dy$, and as a result is a function of doping and V_{GS} . Approximate expressions for this depth have been developed and can be used to provide a correction for the effective oxide thickness as follows:^{197,198}

$$d_m = B_1 \left| Q'_B + \frac{11}{32} Q'_I \right|^{-1/3} \quad (6.9.2)$$

where $B_1 \approx 10^{-9} (\text{C} \cdot \text{cm})^{1/3}$. As can be seen in (6.9.2), d_m and therefore the correction $(\hat{t}_{\text{ox}} - t_{\text{ox}})$ is expected to increase as t_{ox} decreases with device scaling because Q'_B increases as doping increases with scaling. Moreover, the relative oxide thickness increase will accelerate because t_{ox} decreases with scaling.

POLYSILICON GATE DEPLETION. As we have discussed earlier, the standard material for gate formation in modern MOSFET technologies is polycrystalline silicon heavily doped to behave almost like a metal. However, in many technologies it is not possible to dope the polysilicon gate to arbitrarily high concentrations. As a result, depending on the doping type of the gate relative to the channel, the gate may deplete somewhat when an inversion is formed in the channel beneath it.¹⁹⁹⁻²⁰¹ Since it is typical to dope the gate n -type in NMOS and p -type in PMOS in most modern CMOS technologies, this effect is rather common. Gate depletion results in a thickening of the effective oxide thickness, similarly but with different physical origin to that of the d_m effect just discussed. For example, if the gate depletes to a depth of d_p

then the effective oxide thickness increases by $(\epsilon_{\text{ox}}/\epsilon_s)d_p$. Thus combining the two effects the effective oxide thickness becomes

$$\hat{t}_{\text{ox}} = t_{\text{ox}} + \frac{\epsilon_{\text{ox}}}{\epsilon_s} (d_m + d_p) \quad (6.9.3)$$

Similarly to the effect of the finite inversion (or accumulation) thickness, the gate depletion thickness is also bias dependent, but for device modeling purposes it can be assumed constant to some average value.²⁰² For typical technologies circa 1997 the difference $(\hat{t}_{\text{ox}} - t_{\text{ox}})$ is 3 to 5 Å. Even if this stays constant the fractional oxide thickness increase will increase with scaling. For this reason increasing research effort is devoted to replacing the gate material with a pure metal.

QUANTUM MECHANICAL $|V_{T0}|$ INCREASE EFFECT. Another effect of quantum mechanics that also increases with scaling, is a shift $\Delta\psi_s$ in the surface potential required for strong inversion.²⁰² This effect arises from the so-called “energy quantization” of confined particles which precludes electrons or holes from existing at zero energy in the conduction or valence bands. It is a direct consequence of the coupled Poisson-Schrödinger equation solution. Ignoring so-called “bandgap narrowing” effects,^{203,204} an approximate expression can be derived as follows

$$|\Delta\psi_s| = B_2 |Q'_B|^{2/3} + \frac{|Q'_B|}{\epsilon_s} d_m \quad (6.9.4)$$

where $B_2 \approx 500 \text{ V}/(\text{C} \cdot \text{cm}^{-2})^{2/3}$. As can be seen readily in (6.9.4), $\Delta\psi_s$ increases with scaling because $|Q'_B|$ increases with scaling. This surface potential shift manifests itself as an increase in $|V_T|$ which for long-channel devices is

$$|\Delta V_{T0}| = |\Delta\psi_s| + \gamma (\sqrt{\phi_0 + \Delta\psi_s} - \sqrt{\phi_0}) \quad (6.9.5)$$

Again, as can be seen from (6.9.5), $|\Delta V_{T0}|$ increases as devices are scaled down.

TUNNELING THROUGH THE GATE INSULATOR. Constant-field scaling theory requires that gate oxide thickness should scale proportionately to minimum L and W . Given the current trend in MOSFET technology, oxide physical thickness will have to be in the 20-Å regime for gate lengths below 0.13 μm , which are expected around the year 2003.¹⁶⁵ The gate oxide in the MOSFET forms the potential energy barrier that prevents carriers from being injected into the gate from the silicon surface. We have already seen in Sec. 6.6 that energetic—i.e., hot—carriers can be injected into the gate by overcoming the potential barrier. However, when the barrier thickness becomes very small, e.g., 20 Å, then less energetic electrons can tunnel *through* the barrier as another consequence of quantum mechanics.¹⁷⁹ This tunneling effect is expected to limit the scaling of oxide thickness to about 15 Å. Thinning the oxide to much below that will result in the oxide being leaky enough that the insulating property of the MOSFET gate is compromised. A possible solution to this problem can be

the substitution of the silicon dioxide by some other insulating material with significantly higher dielectric constant, so that C'_{ox} , which determines the degree of control of V_{GS} on I_{DS} , can be large without having to make the insulator too thin. Then the scaling-required insulator thickness could be increased relative to silicon dioxide by the ratio of its dielectric constant to that of oxide. This is a subject of intense research at the time of writing of this book.

REFERENCES

1. D. Vandorpe, J. Borel, G. Merckel, and P. Saintot, "An accurate two-dimensional numerical analysis of the MOS transistor," *Solid-State Electronics*, vol. 15, p. 547, 1972.
2. D. P. Kennedy and P. C. Murley, *IBM Journal of Research and Development*, vol. 17, p. 2, 1973.
3. M. S. Mock, "A two-dimensional mathematical model of the insulated gate field-effect transistor," *Solid-State Electronics*, vol. 16, pp. 601-609, 1973.
4. R. Kasai and T. Kimura, "Two-dimensional structure analysis of short-channel C-MOS and transient analysis of the C-MOS circuits," Technical Group, Institute of Electronics and Communications Engineers of Japan, ED76-5, 1976 (in Japanese).
5. J. A. Greenfield and R. W. Dutton, "Nonplanar VLSI device analysis using the solution of Poisson's equation," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 1520-1532, 1980.
6. S. Selberherr, A. Schütz, and H. W. Pötzl, "MINIMOS—A two-dimensional MOS transistor analyzer," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 1540-1549, 1980.
7. B. T. Brown and J. J. H. Miller (editors), *Numerical Analysis of Semiconductor Devices*, Boole Press, Dublin, 1970.
8. E. M. Buturla, P. E. Cottrell, B. M. Grossman, K. A. Salzburg, M. B. Lawlor, and C. T. McMullen, "Three-dimensional finite element simulation of semiconductor devices," *Digest of Technical Papers*, IEEE International Solid-State Circuits Conference, San Francisco, pp. 76-77, 1980.
9. A. Yoshii, H. Kitazawa, M. Tomizawa, S. Horiguchi, and T. Sudo, "A three-dimensional analysis of semiconductor devices," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 184-189, 1982.
10. A. Husain and S. G. Chamberlain, "Three-dimensional simulation of VLSI MOSFET's: the three-dimensional simulation program WATMOS," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 631-638, 1982.
11. W. L. Engl, H. K. Dirks, and B. Meinerzhagen, "Device modeling," *Proceedings of the IEEE*, vol. 71, pp. 10-33, January 1983.
12. M. R. Pinto, C. S. Rafferty, and R. W. Dutton, "PISCES-II: Poisson and continuity equation solver," Stanford Electronics Laboratory Technical Report, September 1984.
13. C. L. Wilson, P. Roitman, and L. Blue, "High accuracy physical modeling of submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 1246-1258, July 1985.
14. C. L. Wilson and J. L. Blue, "Accurate current calculations in two-dimensional MOSFET models," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 2060-2068, October 1985.
15. B. Meinerzhagen, H. K. Dirks, and W. L. Engl, "Quasi-simultaneous solution method: a new highly efficient strategy for numerical MOST simulations," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 2131-2138, October 1985.
16. S. R. Hofstein and F. P. Heinman, "The silicon insulated-gate field-effect transistor," *Proceedings of the IEEE*, vol. 51, pp. 1190-1202, September 1963.
17. C. Goldberg, "Pinch off in insulated-gate field-effect transistors," *Proceedings of the IEEE*, vol. 52, pp. 414-415, April 1964.
18. S. R. Hofstein and G. Warfield, "Carrier mobility and current saturation in the MOS transistor," *IEEE Transactions on Electron Devices*, vol. ED-12, pp. 129-138, March 1965.
19. V. K. G. Reddi and C. T. Sah, "Source to drain resistance beyond pinch-off in Metal-Oxide Semiconductor transistors (MOST)," *IEEE Transactions on Electron Devices*, vol. ED-12, pp. 139-141, March 1965.

20. J. E. Schroeder and R. S. Muller, "IGFET analysis through numerical solution of Poisson's equation," *IEEE Transactions on Electron Devices*, vol. ED-15, pp. 954-961, 1968.
21. H. Shichman and D. A. Hodges, "Modeling and simulation of insulated-gate field-effect transistor switching circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-3, pp. 285-289, September 1968.
22. D. Frohman-Bentchkowsky and A. S. Grove, "Conductance of MOS transistors in saturation," *IEEE Transactions on Electron Devices*, vol. ED-16, pp. 108-113, January 1969.
23. G. Baum and H. Beneking, "Drift velocity saturation in MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-17, pp. 481-482, June 1970.
24. G. Baum, "Driftgeschwindigkeitssättigung bei MOS-Feldeffekttransistoren," *Solid-State Electronics*, vol. 13, pp. 789-798, 1970.
25. R. S. C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley-Interscience, New York, 1970.
26. G. A. Armstrong and J. A. Magowan, "The distribution of mobile carriers in the pinch-off region of an insulated-gate field-effect transistor and its influence on device breakdown," *Solid-State Electronics*, vol. 14, pp. 723-733, 1971.
27. G. Merckel, J. Borel, and N. Z. Cupcea, "An accurate large-signal MOS transistor model for use in computer-aided design," *IEEE Transactions on Electron Devices*, vol. ED-19, pp. 681-690, May 1972.
28. D. Vandorpe, J. Borel, G. Merckel, and P. Saintot, "An accurate two-dimensional numerical analysis of the MOS transistor," *Solid-State Electronics*, vol. 15, pp. 547-557, 1972.
29. A. Popa, "An injection level dependent theory of the MOS transistor in saturation," *IEEE Transactions on Electron Devices*, vol. ED-19, pp. 774-781, 1972.
30. P. Rossel, H. Martinot, and G. Vassilieff, "Accurate two-sections model for MOS transistor in saturation," *Solid-State Electronics*, vol. 19, pp. 51-56, 1976.
31. Y. A. El-Mansy and A. R. Boothroyd, "A simple two-dimensional model for IGFET operation in the saturation region," *IEEE Transactions on Electron Devices*, vol. ED-24, pp. 254-262, 1977.
32. G. Merckel, "CAD models of MOSFETS" in *Process and Device Modelling for Integrated Circuit Design*, F. Van de Wiele, W. L. Engl, and P. G. Jespers (editors), Noordhoff, Leyden, The Netherlands, 1977.
33. F. M. Klaassen, "Review of physical needs for MOS transistors," in *Process and Device Modelling for Integrated Circuit Design*, F. Van de Wiele, W. L. Engl, and P. G. Jespers (editors), Noordhoff, Leyden, The Netherlands, 1977.
34. H. C. Poon, " V_{th} and beyond," presented at the Workshop on Device Modelling for VLSI, Burlingame, California, March 29, 1979; also L. Cong, Bell Laboratories, private communication.
35. B. Hofflinger, H. Sibbert, and G. Zimmer, "Model and performance of hot-electron MOS transistors for VLSI," *IEEE Transactions on Electron Devices*, vol. ED-26, pp. 513-520, April 1979.
36. M. El Nokali and H. Miranda, "A simple model for the MOS transistor in saturation," *Solid-State Electronics*, vol. 29, pp. 591-596, 1986.
37. F. M. Klaassen and W. C. J. de Groot, "Modeling of scaled-down MOS transistors," *Solid-State Electronics*, vol. 23, pp. 237-242, 1980.
38. T. Poorter and J. H. Satter, "A DC model for an MOS transistor in the saturation region," *Solid-State Electronics*, vol. 23, pp. 765-772, 1980.
39. J. R. Brews, "Physics of the MOS transistor," chapter 1 in *Silicon Integrated Circuits, Part A*, D. Kahng (editor), Applied Solid-State Science Series, Academic Press, New York, 1981.
40. S. Liu and L. W. Nagel, "Small-signal MOSFET models for analog circuit design," *IEEE Journal of Solid-State Circuits*, vol. SC-17, pp. 983-998, December 1982.
41. P. Ratnam and C. A. T. Salama, "A new approach to the modeling of nonuniformly doped short-channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 1289-1298, 1984.
42. M. E. Banna, and M. E. Nokali, "A pseudo-two-dimensional analysis of short-channel MOSFETs," *Solid-State Electronics*, vol. 31, pp. 269-274, 1988.
43. P. K. Ko, "Approaches to scaling," pp. 1-37, in *Advanced MOS Device Physics*, N. G. Einspruch and G. Gildenblat (editors), VLSI Electronics, vol. 18, Academic Press, New York, 1989.
44. F. J. Lai and J. Y. Sun, "Analytical one-dimensional model for lightly doped drain (LDD) MOSFET devices," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 2803-2811, 1985.

45. G. S. Huang and C. Y. Wu, "An analytic I-V model for lightly doped drain (LDD) MOSFET devices," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 1311-1321, 1987.
46. C. Turchetti and G. Masetti, "A charge-sheet analysis of short-channel enhancement-mode MOSFETs," *IEEE Journal of Solid-State Circuits*, vol. SC-21, pp. 267-275, 1986.
47. K. Y. Toh, P. K. Ko, and R. G. Meyer, "An engineering model for short-channel MOS devices," *IEEE Journal of Solid-State Circuits*, vol. SC-23, pp. 950-958, 1988.
48. Y. Hu, R. V. H. Booth, and M. H. White, "An analytical model for the lateral channel electric field in LDD structures," *IEEE Transactions on Electron Devices*, vol. ED-37, pp. 2254-2263, 1990.
49. G. T. Cheney and R. A. Kotch, "A simple theory for threshold voltage modulation in IGFETs," *Proceedings of the IEEE*, vol. 56, pp. 837-888, 1968.
50. H. S. Lee, "An analysis of the threshold voltage for short-channel IFGETs," *Solid-State Electronics*, vol. 16, pp. 1407-1414, 1973.
51. R. C. Varschney, "Simple theory for threshold voltage modulation in short-channel MOS transistor," *Electronics Letters*, vol. 9, pp. 600-602, 1973.
52. H. C. Poon, L. D. Yau, R. L. Johnston, and D. Beecham, "D.C. model for short-channel IGFETs," *Technical Digest*, International Electron Devices Meeting, Washington, D.C., pp. 156-159, 1973.
53. L. D. Yau, "A simple theory to predict the threshold voltage of short-channel IGFETs," *Solid-State Electronics*, vol. 17, pp. 1059-1063, 1974.
54. K. O. Jeppson, "Influence of the channel width on the threshold voltage modulation in MOSFETs," *Electronics Letters*, vol. 11, pp. 297-299, July 1975.
55. K. E. Kroell and G. K. Ackermann, "Threshold voltage of narrow channel field effect transistors," *Solid-State Electronics*, vol. 19, pp. 77-81, 1976.
56. W. P. Noble and P. E. Cottrell, "Narrow width effects in insulated gate field effect transistors," *Technical Digest*, International Electron Devices Meeting, pp. 582-586, 1976.
57. W. R. Bandy and D. P. Kokalis, "A simple approach for accurately modeling the threshold voltage of short-channel MOST's," *Solid-State Electronics*, vol. 20, pp. 675-680, 1977.
58. D. J. Coe, H. E. Brakman, and K. H. Nicholas, "A simple approach for accurately modeling the threshold voltage of short-channel MOST's," *Solid-State Electronics*, vol. 20, p. 993, 1977.
59. G. Merckel, "Short channels—scaled down MOSFETs," in *Process and Device Modelling for Integrated Circuit Design*, F. Van de Wiele, W. L. Engl, and P. G. Jespers (editors), Noordhoff, Leyden, The Netherlands, 1977.
60. R. R. Troutman and A. G. Fortino, "Simple model for threshold voltage in short-channel IGFETs," *IEEE Transactions on Electron Devices*, vol. ED-24, pp. 1266-1268, October 1977.
61. G. W. Taylor, "Subthreshold conduction in MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-25, pp. 337-350, March 1978.
62. P. P. Wang, "Device characteristics of short-channel and narrow width MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-25, pp. 779-786, 1978.
63. E. Sun, "Short-channel MOS modeling for CAD," *Proceedings of the Twelfth Annual Asilomar Conference on Circuits, Systems, and Computers*, Pacific Grove, CA, pp. 493-499, November 1978.
64. W. Fichter and H. W. Potzl, "MOS modeling by analytical approximations. I. Subthreshold current and threshold voltage," *International Journal of Electronics*, vol. 46, pp. 33-55, 1979.
65. L. M. Dang, "A simple current model for short-channel IGFET and its application to circuit simulation," *IEEE Journal of Solid-State Circuits*, vol. SC-14, pp. 358-367, April 1979.
66. R. R. Troutman, "VLSI limitations from drain-induced barrier lowering," *IEEE Journal of Solid-State Circuits*, vol. SC-14, pp. 383-391, April 1979.
67. G. W. Taylor, "The effects of two-dimensional charge sharing on the above-threshold characteristics of short-channel IGFETs," *Solid-State Electronics*, vol. 22, pp. 701-717, 1979.
68. T. Toyabe and S. Asai, "Analytical models of threshold voltage and breakdown voltage of short-channel MOSFET's derived from two-dimensional analysis," *IEEE Transactions on Electron Devices*, vol. ED-26, pp. 453-461, 1979.
69. H. Masuda, M. Makai, and M. Kubo, "Characteristics and limitation of scaled-down MOSFET's due to two-dimensional field effect," *IEEE Transactions on Electron Devices*, vol. ED-26, pp. 980-986, 1979.

70. G. Merckel, "A simple model of the threshold voltage of short and narrow channel IGFETs," *Solid-State Electronics*, vol. 23, pp. 1207-1213, 1980.
71. A. Vladimirescu and S. Liu, "The simulation of MOS integrated circuits using SPICE 2," *Memo-randum No. UCB/ERL M80/7*, Electronics Research Laboratory, University of California, Berkeley, February 1980.
72. P. K. Chatterjee and J. E. Leiss, "An analytic charge-sharing predictor model for sub-micron MOS-FETs," *Technical Digest*, IEEE International Electron Devices Meeting, Washington, D.C., pp. 28-33, 1980.
73. K. N. Ratnakumar, J. D. Meindl, and D. L. Scharfetter, "New IGFET short channel threshold volt-age model," *Technical Digest*, IEEE International Electron Devices Meeting, Washington, D.C., pp. 204-206, 1981.
74. P. P. Guebels and F. Van de Wiele, "A charge sheet model for small geometry MOSFETs," *Techni-cal Digest*, IEEE International Electron Devices Meeting, Washington, D.C., pp. 211-214, 1981.
75. T. N. Nguyen and J. D. Plummer, "Physical mechanisms responsible for short channel effects in MOS devices," *Technical Digest*, IEEE International Electron Devices Meeting, Washington, D.C., pp. 596-599, 1981.
76. A. A. Naem and A. R. Boothroyd, "Compensation tendency of short-channel and narrow-channel effects in small-geometry IGFETs," *Electronics Letters*, vol. 18, pp. 135-136, February 4, 1982.
77. R. Kasai, K. Yokoyama, A. Yoshii, and T. Sudo, "Threshold-voltage analysis of short- and narrow-channel MOSFETs by three-dimensional computer simulation," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 870-876, March 1982.
78. L. A. Akers and C. S. Chao, "A closed-form threshold voltage expression for a small-geometry MOSFET," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 776-778, April 1982.
79. L. A. Akers and J. J. Sanchez, "Threshold voltage models of short, narrow, and small geometry MOSFET's: a review," *Solid-State Electronics*, vol. 25, pp. 621-641, 1982.
80. P. P. Guebels and F. Van de Wiele, "A small geometry MOSFET model for CAD applications," *Solid-State Electronics*, vol. 26, pp. 267-273, 1983.
81. C. R. Ji and C. T. Sah, "Analysis of the narrow gate effect in submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, vol. ED-30, pp. 1672-1677, December 1983.
82. C. R. Viswanathan, B. C. Burkey, G. Lubberts, and T. J. Tredwell, "Threshold voltage in short-chan-nel MOS devices," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 932-940, May 1985.
83. A. M. Asenov, E. N. Stefanov, B. Z. Antov, and P. K. Vitanov, "Numerical analysis of MOS tran-sistor effective channel width," *Electronics Letters*, vol. 21, pp. 595-597, July 4, 1985.
84. T. W. Tang, Q. L. Zhang, and D. H. Navon, "Analytical model for predicting threshold voltage in submicrometer-channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 1890-1895, September 1985.
85. C. T. Wang, "A threshold voltage expression for small-size MOSFET's based on an approximate three-dimensional analysis," *IEEE Transactions on Electron Devices*, vol. ED-23, pp. 160-164, January 1986.
86. S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.
87. F. M. Klaassen, "MOS device modelling," in *Design of MOS VLSI Circuits for Telecommunica-tions*, Y. Tsividis and P. Antognetti (editors), Prentice-Hall, Englewood Cliffs, N.J., 1985.
88. N. D. Arora, *MOSFET Models for VLSI Circuit Simulation—Theory and Practice*, Computational Microelectronics Series, S. Selberherr (editor), Springer-Verlag, Vienna and New York, 1993.
89. *MEDICI User's Manual*, Technology Modeling Associates, Santa Clara, Calif., 1996.
90. G. W. Taylor, "Subthreshold conduction in MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-25, pp. 337-350, 1978.
91. M. El Banna and M. El Nokali, "A pseudo-two-dimensional analysis of short-channel MOSFETs," *Solid-State Electronics*, vol. 31, pp. 269-274, 1988.
92. J. D. Kendall and A. R. Boothroyd, "A two-dimensional analytical solution of the Poisson and cur-rent continuity equations for the short-channel MOSFET," *Solid-State Electronics*, vol. 33, pp. 537-551, 1990.
93. M. Conti and C. Turchetti, "On the short-channel theory for MOS transistor," *IEEE Transactions on Electron Devices*, vol. ED-38, pp. 2657-2663, 1991.

94. T. A. Fjeldy and M. Shur, "Threshold voltage modeling and the subthreshold regime of operation of short-channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. 40, pp. 137-145, 1993.
95. Z-H Liu, C. Hu, J-H Huang, T-Y Chan, M-C Jeng, P. K. Ko, and Y. C. Cheng, "Threshold voltage model for deep-submicrometer MOSFETs," *IEEE Transactions on Electron Devices*, vol. 40, pp. 86-95, 1993.
96. B. Iniguez, "Comments on 'Threshold voltage model for deep-submicrometer MOSFETs,'" *IEEE Transactions on Electron Devices*, vol. 42, p. 1712, 1995.
97. A. Kloss and A. Kostka, "A new physics-based, predictive compact model for small geometry MOSFETs including two-dimensional calculations with a close link to process and layout data," *IEDM Technical Digest*, pp. 147-150, 1996.
98. N. D. Arora, R. Rios, C-L Huang, and K. Raol, "PCIM: A physically based continuous short-channel IGFET model for circuit simulation," *IEEE Transactions on Electron Devices*, vol. 41, pp. 988-997, 1994.
99. N. D. Arora and M. Sharma, "Modeling the anomalous threshold voltage behavior of submicron MOSFETs," *IEEE Electron Device Letters*, vol. EDL-13, pp. 92-94, 1992.
100. C. Rafferty, H-H. Vuong, S. Eshraghi, M. Giles, M. Pinto, and S. Hilenius, "Explanation of reverse short-channel effect by defect gradients," *IEDM Technical Digest*, pp. 311-314, 1993.
101. A. G. Lewis and J. Y. Chen, "Current trends in MOS Process Integration," pp. 40-117, in *Advanced MOS Device Physics*, N. G. Einspruch and G. Gildenblat (editors), VLSI Electronics, vol. 18, Academic Press, New York, 1989.
102. A. Kurosawa, T. Shibata, and H. Iozuka, "A new bird's beak free isolation technology for VLSI devices," *IEDM Technical Digest*, pp. 384-387, 1981.
103. L. A. Akers, M. Sugino, and J. M. Ford, "Characterization of the inverse-narrow-width effect," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 2476-2484, 1987.
104. K. K. Hsueh, J. J. Sanchez, T. A. Demassa, and L. A. Akers, "Inverse-narrow-width effects and small-geometry MOSFET threshold voltage model," *IEEE Transactions on Electron Devices*, vol. ED-36, pp. 325-338, 1988.
105. K. Ohe, S. Odanaka, K. Moriyama, T. Hora, and G. Fuse, "Narrow-width effects of shallow trench-isolated CMOS with n⁺-polysilicon gate," *IEEE Transactions on Electron Devices*, vol. ED-36, pp. 1110-1115, 1989.
106. R. C. Vankemmel and K. M. De Meyer, "A study of the corner effects in trench-like isolated structures," *IEEE Transactions on Electron Devices*, vol. ED-37, pp. 168-175, 1990.
107. P. Richman, *MOSFET's and Integrated Circuits*, John Wiley, New York, 1973.
108. R. A. Stuart and W. Eccleston, "Punchthrough currents in short-channel M.O.S.T. devices," *Electronics Letters*, vol. 9, pp. 586-588, December 1973.
109. J. R. Brews, "Geometrical factors in avalanche punchthrough erase," *IEEE Transactions on Electron Devices*, vol. ED-24, pp. 1108-1116, August 1977.
110. J. J. Barnes, K. Shimohigashi, and R. W. Dutton, "Short-channel MOSFETs in the punchthrough current mode," *IEEE Transactions on Electron Devices*, vol. ED-26, pp. 446-453, April 1979.
111. F. S. Hsu, R. S. Muller, C. Hu, and P-K. Ko, "A simple punchthrough model for short-channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-30, pp. 1354-1359, October 1983.
112. F. N. Trofimenkoff, "Field-dependent mobility analysis to the field-effect transistor," *Proceedings of the IEEE*, vol. 53, pp. 1765-1766, January 1965.
113. R. H. Crawford, *MOSFET in Circuit Design*, McGraw-Hill, New York, 1967.
114. B. Hoeneisen and C. A. Mead, "Current-voltage characteristics of small size MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-19, pp. 382-383, 1972.
115. P. Smith, M. Inoue, and J. Frey, "Electron velocity in Si and GaAs at very high electric fields," *Applied Physics Letters*, vol. 37, pp. 797-798, 1980.
116. K. K. Thornber, "Relation of drift velocity to low-field mobility and high-field saturation velocity," *Journal of Applied Physics*, vol. 51, pp. 2127-2133, April 1980.
117. M. H. White, F. Van de Wiele, and J. P. Lambot, "High-accuracy MOS models for computer-aided design," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 899-906, May 1980.
118. C. G. Sodini, P-K. Ko, and J. L. Moll, "The effect of high fields on MOS device and circuit performance," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 1386-1393, October 1984.

119. J. G. Ruch, "Electron dynamics in short-channel field-effect transistors," *IEEE Transactions on Electron Devices*, vol. ED-19, pp. 652-654, May 1972.
120. T. Kobayashi and K. Saito, "Two-dimensional analysis of velocity overshoot effects in ultrashort-channel Si MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 788-792, April 1985.
121. G. G. Shahidi, D. A. Antoniadis, and H. I. Smith, "Electron velocity overshoot at room and liquid nitrogen temperatures in silicon inversion layers," *IEEE Electronic Device Letters*, vol. EDL-9, pp. 94-96, 1988.
122. R. Coen and R. S. Muller, "Velocity of surface carriers in inversion layers of silicon," *Solid-State Electronics*, vol. 23, pp. 35-40, 1980.
123. W. W. Lattin and J. L. Rutledge, "Impact ionization current in MOS devices," *Solid-State Electronics*, vol. 16, p. 1043, 1973.
124. T. Kamata, K. Tanabashi, and K. Kobayashi, "Substrate current due to impact ionization in MOSFETs," *Japanese Journal of Applied Physics*, vol. 15, p. 1127, 1976.
125. E. Sun, J. Moll, J. Berger, and B. Alders, "Breakdown mechanism in short-channel MOS transistors," *Technical Digest*, IEEE International Electron Devices Meeting, Washington, D.C., pp. 478-482, 1978.
126. J. Matsunaga, M. Konaka, S. Kohyama, and H. Iizuku, "Design limitations due to substrate currents and secondary impact ionization electrons in NMOS LSI's," *Proceedings of the Eleventh International Conference on Solid-State Devices*, Tokyo, p. 45, August 1979.
127. P. K. Chatterjee, "VLSI dynamic NMOS design constraints due to drain induced primary and secondary impact ionization," *Technical Digest*, IEEE International Electron Devices Meeting, Washington, D.C., pp. 14-17, 1979.
128. B. Eitan and D. Frohman-Bentchkowsky, "Holding time degradation in dynamic MOS RAM by injection-induced electron currents," *IEEE Transactions on Electron Devices*, vol. ED-28, pp. 1515-1519, December 1981.
129. Y. El-Mansy, "MOS device and technology constraints in VLSI," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 567-573, April 1982.
130. S. Tam, P. Ko, F. C. Hsu, C. Hu, and P. S. Muller, "Hot electron-induced excess currents in n-channel MOSFETs," Device Research Conference, Colorado State University, Fort Collins, Colo., June 1982.
131. F.-C. Hsu, P.-K. Ko, S. Tam, C. Hu, and R. S. Muller, "An analytical breakdown model for short-channel MOSFET's," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 1735-1740, November 1982.
132. W. Muller, L. Risch, and A. Schutz, "Short-channel MOS transistors in the avalanche-multiplication regime," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 1778-1784, November 1982.
133. R. Kuhnert, C. Werner, and A. Schutz, "A novel impact-ionization model for 1- μ m MOSFET simulation," *IEEE Transactions on Electron Devices*, vol. ED-32, no. 6, pp. 1057-1063, June 1985.
134. H. Hana, Y. Okamoto, and H. Ohnuma, *Japanese Journal of Applied Physics*, vol. 9, p. 1103, 1970.
135. T. H. Ning, C. M. Osburn, and H. N. Yu, "Effect of electron tapping on IGFET characteristics," *Journal of Electronics Materials*, vol. 6, pp. 65-76, 1977.
136. T. H. Ning, P. W. Cook, R. H. Dennard, C. M. Osburn, S. E. Shuster, and H. N. Yu, "1- μ m MOSFET VLSI technology: Part IV—Hot-electron design constraints," *IEEE Transactions on Electron Devices*, vol. ED-26, pp. 346-353, April 1979.
137. P. E. Cottrell, R. R. Troutman, and T. H. Ning, "Hot electron emission in n-channel IGFETs," *IEEE Transactions on Electron Devices*, vol. ED-26, pp. 520-533, April 1979.
138. R. B. Fair and R. C. Sun, "Threshold voltage instability in MOSFETs due to channel hot-hole emission," *IEEE Transactions on Electron Devices*, vol. ED-28, pp. 83-94, January 1981.
139. S. Tam, F.-C. Hsu, C. Hu, R. S. Muller, and P. K. Ko, "Hot-electron currents in very short channel MOSFET's," *IEEE Electronic Device Letters*, vol. EDL-4, pp. 249-251, July 1983.
140. C. Hu, S. C. Tam, F. C. Hsu, P. K. Ko, T. Y. Chan, and K. W. Terrill, "Hot-electron induced MOSFET degradation—model, monitor, and improvement," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 375-385, 1985.
141. E. Takeda, "Hot-carrier effects in submicrometer MOS VLSI's," *IEEE Proceedings*, vol. 131, part I, pp. 153-162, October 1984.

142. T. Tsuchiya and J. Frey, "Relationship between hot-electrons/holes and degradation of p- and n-channel MOSFET's," *IEEE Electron Device Letters*, vol. EDL-6, pp. 8-11, January 1985.
143. K. R. Hofmann, C. Werner, W. Weber, and G. Dorda, "Hot-electron and hole emission effects in short n-channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 691-699, March 1985.
144. Y. El-Mansy, "MOS device and technology constraints in VLSI," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 567-573, April 1982.
145. K. Brennan and K. Hess, "A theory of enhanced impact ionization due to the gate field and mobility degradation in the inversion layer of MOSFETs," *IEEE Electron Device Letters*, vol. EDL-7, pp. 86-88, February 1986.
146. W. Müller, L. Risch, and A. Schütz, "Short-Channel MOS Transistors in the Avalanche-Multiplication Regime," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 1778-1784, 1982.
147. J. Mar, S. S. Li, and S. Y. Yu, "Substrate Current Modelling for Circuit Simulation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits*, vol. CAD-1, pp. 183-186, 1982.
148. H. C. de Graaff and F. M. Klaassen, "Compact Transistor Modeling for Circuit Design," Springer-Verlag, Vienna and New York, 1990.
149. M. Fukuma and W. Lui, "MOSFET substrate current model including energy transport," *IEEE Electron Devices Letters*, vol. EDL-8, pp. 214-216, 1987.
150. J. W. Slotboom, G. Streutker, G. J. T. Devids, and P. B. Hartog, "Surface impact ionization in silicon devices," *IEDM Technical Digest*, pp. 494-497, 1987.
151. C. T. Wang, "An improved hot-electron-emission model for simulating the gate-current characteristic of MOSFETs," *Solid-State Electronics*, vol. 31, pp. 229-231, 1988.
152. J. Higman, I. C. Kizilyalli, and K. Hess, "Nonlocality of the electron ionization coefficient in n-MOSFETs: An analytical approach," *IEEE Electron Devices Letters*, vol. EDL-9, pp. 399-401, 1988.
153. N. Goldsman and J. Frey, "Electron energy distribution for calculation of gate leakage current in MOSFETs," *Solid-State Electronics*, vol. 31, pp. 1089-1092, 1988.
154. B. Meinerzhagen, "Consistent gate and substrate current modeling based on energy transport and the lucky electron concept," *IEDM Technical Digest*, pp. 504-507, 1988.
155. C. G. Hwang and R. W. Dutton, "Substrate current model for submicrometer MOSFETs based on mean free path analysis," *IEEE Transactions on Electron Devices*, vol. ED-36, pp. 1348-1354, 1989.
156. G-S Huang and C-Y Wu, "An analytic saturation model for drain and substrate current of conventional and LDD MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-37, pp. 1667-1677, 1990.
157. Y. A. El-Mansy and D. M. Caughey, "Modeling weak avalanche multiplication currents in IGFETs and SOS transistors for CAD," *IEDM Technical Digest*, pp. 31-34, 1975.
158. C. Hu, "Hot-carrier effects," pp. 119-160, in *Advanced MOS Device Physics*, N. G. Einspruch and G. Gildenblat (editors), VLSI Electronics, vol. 18, Academic Press, New York, 1989.
159. S. Ogura, P. J. Tsang, W. W. Walker, D. L. Critchlow, and J. F. Shepard, "Design and characteristics of the lightly doped drain-source (LDD) insulated-gate field-effect transistor," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 1359-1366, August 1980.
160. E. Takeda, H. Kume, Y. Nakagome, T. Makino, A. Shimizu, and S. Asai, "An As-P (n^+n^-) double diffused rain MOSFET for VLSI's," *IEEE Transactions on Electron Devices*, vol. ED-30, pp. 652-657, 1983.
161. H. Mikoshiba, T. Horiuchi, and K. Hamano, "Comparison of drain structures in n-channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 140-144, January 1986.
162. J. J. Sanchez, K. K. Hsueh, and T. A. DeMassa, "Drain-engineered hot-electron-resistant device structures—A review," *IEEE Transactions on Electron Devices*, vol. ED-36, pp. 1125-1131, 1989.
163. L. C. Parillo, "VLSI process integration," in *VLSI Technology*, S. M. Sze (editor), McGraw-Hill, New York, 1983.
164. S. Wolf, "Silicon processing in the VLSI era," *The Submicron MOSFET*, vol. 3, Lattice Press, Sunset Beach, Calif., 1995.

165. "The national technology roadmap for semiconductors: Technology Needs," Semiconductor Industry Association, San Jose, Calif., 1997.
166. R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. SC-9, pp. 256-268, 1974.
167. H.-N. Yu, A. Reisman, C. M. Osborn, and D. L. Critchlow, "1 μm MOSFET VLSI technology, Part I: An Overview," *IEEE Transactions on Electron Devices*, vol. ED-26, pp. 318-324, April 1979.
168. R. H. Dennard, F. H. Gaensslen, E. J. Walker, and P. W. Cook, "1 μm MOSFET VLSI technology: Part II—Device designs and characteristics for high-performance logic applications," *IEEE Transactions on Electron Devices*, vol. ED-26, pp. 325-333, April 1979.
169. J. R. Brews, W. Fichtner, E. H. Nicolian, and S. M. Sze, "Generalized guide for MOSFET miniaturization," *IEEE Electron Device Letters*, vol. EDL-1, pp. 2-3, January 1980.
170. Y. El-Mansy, "MOS device and technology constraints in VLSI," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 567-573, April 1982.
171. VLSI Laboratory, Texas Instruments, "Technology and design challenges of MOS VLSI," *IEEE Journal of Solid-State Circuits*, vol. SC-17, pp. 442-448, June 1982.
172. G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to a 1/4 micron MOSFET design," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 452-462, April 1984.
173. A. Reisman, "Device, circuit, and technology scaling to micron and submicron dimensions," *Proceedings of the IEEE*, vol. 71, pp. 550-565, May 1983.
174. J. H. King, "A novel approach to silicon gate CMOS device scaling," *Solid-State Electronics*, vol. 26, pp. 879-891, 1983.
175. E. Sangiorgi, E. A. Hofstatter, R. K. Smith, P. F. Bechtold, and W. Fichtner, "Scaling issues related to high field phenomena in submicrometer MOSFETs," *IEEE Electron Device Letters*, vol. EDL-7, pp. 115-118, February 1986.
176. P. K. Chatterjee, W. R. Hunter, T. C. Holloway, and Y. T. Lin, "The impact of scaling laws on the choice of n-channel or p-channel for MOS VLSI," *Electron Device Letters*, vol. EDL-1, pp. 220-223, October 1980.
177. Y. Tsividis, "Moderate inversion in MOS devices," *Solid-State Electronics*, vol. 25, pp. 1099-1104, 1982; see also Erratum, *ibid.*, vol. 26, p. 823, 1983.
178. G. Baccarani and M. R. Wordeman, "Transconductance degradation in thin-oxide MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-30, pp. 1295-1304, October 1983.
179. B. Hoeneisen and C. A. Mead, "Fundamental limitations in microelectronics—I. MOS technology," *Solid-State Electronics*, vol. 15, pp. 819-829, 1972.
180. R. M. Swanson and J. D. Meindl, "Fundamental performance limits of MOS integrated circuits," *Digest of Technical Papers*, International Solid-State Circuits Conference, Philadelphia, pp. 110-111, February 1975.
181. R. W. Keys, "Physical limits in digital electronics," *Proceedings of the IEEE*, vol. 63, pp. 740-767, May 1975.
182. O. G. Folberth and J. H. Bleher, "The fundamental limitations of digital semiconductor technology," *Microelectronics Journal*, vol. 9, pp. 33-41, 1979.
183. K. N. Ratnakumar and J. D. Meindl, "Performance limits of E/D NMOS VLSI," *Digest of Technical Papers*, International Solid-State Circuits Conference, San Francisco, pp. 72-73, February 1980.
184. J. Meindl, "Circuit scaling limits for ultra large scale integration," *Digest of Technical Papers*, International Solid-State Circuits Conference, New York, pp. 36-37, February 1981.
185. C. Svensson, "VLSI physics," *Integration*, vol. 1, pp. 3-19, 1983.
186. H. Shichijo, "A re-examination of practical performance limits of scaled n-channel and p-channel MOS devices for VLSI," *Solid-State Electronics*, vol. 26, pp. 969-986, 1983.
187. J. Pfister, J. D. Shott, and J. D. Meindl, "Performance limits of NMOS and CMOS," *Digest of Technical Papers*, International Solid-State Circuits Conference, San Francisco, pp. 158-159, February 1984.
188. E. Takeda, G. A. C. Jones, and H. Ahmed, "Constraints on the application of 0.5 μm MOSFET's to VLSI systems," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 322-327, February 1985.

189. J. R. Pfister, J. D. Shott, and J. D. Meindl, "Performance limits of CMOS ULSI," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 333–343, February 1985.
190. P. Antognetti, C. Lombardi, and D. A. Antoniadis, "Use of 2-D MOS simulation in the study of doping profile influence on S/D resistance in short-channel MOSFETs," *IEDM Technical Digest*, pp. 574–577, 1981.
191. G. Baccarani and G. A. Sai-Halasz, "Spreading resistance in submicron MOSFETs," *IEEE Electron Device Letters*, vol. EDL-4, pp. 27–29, February 1983.
192. K. K. Ng, R. J. Bayruns, and S. C. Fang, "The spreading resistance of MOSFETs," *IEEE Electron Device Letters*, vol. EDL-6, pp. 195–198, April 1985.
193. K. K. Ng and W. T. Lynch, "Analysis of the gate-voltage-dependent series resistance of MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 965–972, 1986.
194. K. K. Ng and W. T. Lynch, "The impact of series resistance on MOSFET scaling," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 503–509, 1987.
195. C. J. Koeneke and W. T. Lynch, "Lightly doped Schottky MOSFET," *Technical Digest, International Electron Devices Meeting*, San Francisco, p. 466, 1982.
196. F. Stern and W. E. Howard, "Properties of semiconductor surface inversion layers in the electric quantum limit," *Physical Review B*, vol. 163, pp. 816–835, 1967.
197. F. Stern, "Quantum properties of surface space-charge layers," *CRC Critical Reviews in Solid-State Sciences*, pp. 499–514, 1974.
198. R. Rios and N. D. Arora, "Determination of ultra-thin gate oxide thickness for CMOS structures using quantum effects," *IEDM Technical Digest*, pp. 613–616, 1994.
199. C.-Y. Lu, M. Sung, H. C. Kirsch, S. T. Hillenius, T. E. Smith, and L. Manchanda, "Anomalous C-V characteristics of implanted poly MOS structure in n⁺/p⁺ dual-gate CMOS technology," *IEEE Electron Device Letters*, vol. 10, pp. 192–194, 1989.
200. N. D. Arora, R. Rios, and C.-L. Huang, "Modeling the polysilicon depletion effect and its impact on submicrometer CMOS circuit performance," *IEEE Transactions on Electron Devices*, vol. 42, pp. 935–943, May 1995.
201. B. Riccò, R. Versari, and D. Esseni, "Characterization of polysilicon-gate depletion in MOS structures," *IEEE Electron Device Letters*, vol. 17, pp. 103–105, March 1996.
202. C. Y. Wong, J. Y.-C. Sun, Y. Taur, C. S. Oh, R. Angelucci, and B. Davari, "Doping of n⁺ and p⁺ polysilicon in dual-gate CMOS process," *IEDM Technical Digest*, pp. 238–241, 1988.
203. M. J. van Dort, P. H. Woerlee, A. J. Walker, C. A. H. Juffermans, and H. Lifka, "Influence of high substrate doping levels on the threshold voltage and the mobility of deep submicrometer MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-39, pp. 932–938, 1994.
204. J. W. Slotboom and H. C. de Graaf, "Measurements of bandgap narrowing in Si bipolar transistors," *Solid-State Electronics*, vol. 19, pp. 857–862, 1976.

PROBLEMS

- 6.1. Derive (6.2.1) using (1.2.13) (see Appendix B).
- 6.2. Prove (6.2.12) and state the condition(s) for its validity.
- 6.3. Prove (6.2.18) and show that, in the limit of V_A approaching infinity, \hat{V}_{DS} approaches V'_{DS} . Investigate how close \hat{V}_{DS} is to V'_{DS} for $\alpha = 1$, V_A in the range of 20 to 50 V, and $V_{GS} - V_T$ in the range of 1 to 5 V.
- 6.4. Consider a model consisting of (4.5.31) in nonsaturation, and (6.2.5) and (6.2.6) in saturation. Assume the two regions are adjacent at the point $V_{DS} = \hat{V}_{DS}$. Find the value of \hat{V}_{DS} that will guarantee continuity of $I_{DS}(V_{DS})$ and of its slope.
- 6.5. Consider a transistor with $N_A = 4 \times 10^{17} \text{ cm}^{-3}$, $t_{ox} = 60 \text{ \AA}$, $V_{FB} = -0.9 \text{ V}$, and junction depth of $0.3 \text{ }\mu\text{m}$, biased at $V_{GS} = 2 \text{ V}$, $V_{SB} = 0$. Consider channel length modulation as described by (6.2.8). Plot l_p vs. V_{DS} for values of V_{DS} between $V'_{DS} + 1 \text{ V}$ and 3 V . How long should the device channel be to ensure a l_p/L of 1 percent or less for the voltage range considered?

- 6.6. Examine carefully the arguments related to charge sharing up to (6.3.2). Identify the points where arbitrary assumptions were used.
- 6.7. Prove (6.3.6) and show that it reduces to (6.3.7) if d_B/d_j is small.
- 6.8. Equation (6.3.6) was derived for L large enough so that a trapezoidal region could be defined as shown in Fig. 6.7a. Investigate the case where L is so small that the trapezoid becomes a triangle. Derive expressions for \hat{Q}_B/Q_B , \hat{V}_T , and ΔV_{TL} in this case. Assume punchthrough does not occur.
- 6.9. Compare the model of (6.3.2b), (6.3.5), and (6.3.6) to the model of (6.3.9), (adjusting β_1 if necessary for best fit) for the process of Prob. 6.5 and with $d_j = 0.8 \mu\text{m}$ by plotting \hat{V}_T vs. V_{SB} for $L = 2, 1$, and $0.5 \mu\text{m}$.
- 6.10. Consider the device of Prob. 6.5; assume $V_{DS} = 0$. Plot \hat{V}_T vs. L for L between 0.5 and $10 \mu\text{m}$, and for $V_{SB} = 0, 1$, and 3 V , using (6.3.9).
- 6.11. Proceeding along the lines of the derivation of (6.3.6), derive a similar expression for the case of a relatively small nonzero V_{DS} , assuming that the edge of the depletion region in the center part of the device is horizontal at a depth equal to the average value of the source and drain depletion regions. From this find \hat{V}_T and ΔV_{TL} . Show the details of the derivation of (6.3.11) to (6.3.15).
- 6.12. Give a plot like the one in Fig. 6.9, for the process of Prob. 6.5, V_{DS} between 0 and 3 V , and $L = 10 \mu\text{m}, 2 \mu\text{m}, 1 \mu\text{m}$ and $0.5 \mu\text{m}$.
- 6.13. Prove (6.3.22).
- 6.14. Equation (6.3.23) was derived for $V_{DS} = 0$. Attempt to extend this result to the case of a small nonzero V_{DS} . (Note: The author is not aware of any experimental evidence of \hat{V}_T dependence on V_{DS} for narrow-channel devices; the result obtained in this problem should be taken with a grain of salt.)
- 6.15. For the device of Prob. 6.5, and assuming LOCOS isolation and that the channel is long, plot V_T vs. W for W between 0.5 and $5 \mu\text{m}$ and for $V_{SB} = 0, 1$, and 2 V ($V_{DS} = 0$).
- 6.16. Consider a device with LOCOS isolation and with a long narrow channel, in which the effects of effective threshold increase and effective channel narrowing are both evident (Sec. 6.3.3). Suggest a measuring technique that can be used to determine ΔV_{TW} and $\Delta \hat{W}$.
- 6.17. Define an effective depletion region charge \hat{Q}_B (Sec. 6.3) in the case of channels that are *both* short *and* narrow. Give the resulting expression for \hat{V}_T in terms of \hat{Q}_B/Q_B , and attempt to justify a common approach for combining short- and narrow-channel effects when each effect acting by itself is small, which assumes that $V_T(V_{DS}) = V_T + \Delta V_{TL}(L, V_{DS}) + \Delta V_{TW}(W, V_{DS})$, where ΔV_{TL} and ΔV_{TW} are modeled as in Sec. 6.3. Assume LOCOS isolation.
- 6.18. Develop a formula for roughly estimating the value of V_{DS} at which punchthrough begins.
- 6.19. Derive expressions corresponding to (6.5.9) and (6.5.10) assuming, instead of (6.5.3), a two-section piecewise-linear model for $v_d(\mathcal{E}_x)$ (i.e., assume that $|v_d| = \mu|\mathcal{E}_x|$ for $|\mathcal{E}_x| \leq \mathcal{E}_c$, and $|v_d| = |v_d|_{\text{max}}$ for $|\mathcal{E}_x| > \mathcal{E}_c$).
- 6.20. Prove (6.5.12) and show that, in the absence of velocity saturation effects ($L\mathcal{E}_c$ approaching infinity), V'_{DS} as given there reduces to $(V_{GS} - V_T)/\alpha$.

- 6.21.** Give the complete equations for nonsaturation and saturation in the presence of velocity saturation effects, using a model based on (6.5.11), (6.5.5), and (6.5.6). The value of V'_{DS} should be slightly modified from (6.5.12) to ensure continuity of both I_{DS} and dI_{DS}/dV_{DS} at $V_{DS} = V'_{DS}$.
- 6.22.** (a) Consider a device with $\mu C'_{ox} = 100 \mu\text{A}/\text{V}^2$, $\alpha = 1.2$, $\mathcal{E}_c = 2.5 \times 10^4 \text{ V/cm}$, and $V_T = 0.5 \text{ V}$. Neglect channel length modulation for simplicity. By using the model of Example 6.2, plot I_{DS} vs. V_{DS} for V_{DS} up to 5 V, with V_{GS} as a parameter ($V_{GS} = 1, 2, 3 \text{ V}$), and for $W = L = 5, 2$, and $1 \mu\text{m}$.
- (b) Repeat for the above values of L , but keep W constant at $5 \mu\text{m}$. Discuss the results obtained. Neglect the effect of channel length and width on the threshold voltage.
- 6.23.** Extend Table 6.2, giving values for all entries in Table 6.1, for the cases of constant voltage, quasi-constant voltage, and generalized scaling.

CHAPTER 7

THE MOS TRANSISTOR IN DYNAMIC OPERATION— LARGE-SIGNAL MODELING

7.1 INTRODUCTION

The MOS transistor has been treated in previous chapters with the assumption that all terminal voltages are constant. However, the device is usually employed in circuits with *varying* terminal voltages. Such “dynamic” operation causes the transistor charges to vary, and the charge changes must be supplied from the outside world by extra currents flowing through the device terminals; here “extra” refers to currents not predicted by dc theory.

The subject of this chapter is the evaluation of charges and terminal currents under dynamic operation, without placing restrictions on the magnitude of the variations; i.e., we will deal with the *large-signal* dynamic operation of the MOS transistor.¹⁻⁷¹† We will concentrate on the device part between source and drain, containing

†The references include some papers mainly on small-signal operation, which, nevertheless, contain material pertinent to our discussion in this chapter. Chapters 8 and 9 contain an extensive discussion of small-signal operation and many more related references.

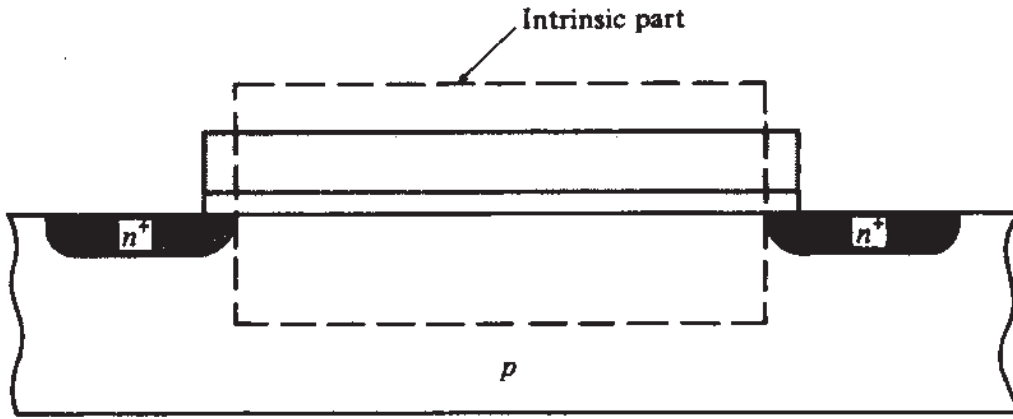


FIGURE 7.1
Indicating the intrinsic part of a transistor.

the inversion layer, the depletion region, the oxide, and the gate. This part is shown enclosed in a broken line in Fig. 7.1. It is called the *intrinsic* part and is the part mainly responsible for transistor action. The rest of the device constitutes the *extrinsic* part and is responsible for parasitic effects, which can limit overall performance. We will postpone discussion of the extrinsic device part until Chap. 8.

Unless stated otherwise in this chapter we will assume long and wide channels and a uniform substrate, and we will ignore the effects of the resistances of the substrate and the gate. Also, we will assume that no hot electron effects are present; the gate and substrate currents in this chapter are not due to such effects, but rather to changes in the gate and substrate charges, respectively.

7.2 QUASI-STATIC OPERATION

We consider a fictitious device as shown in Fig. 7.2, with source and drain omitted to emphasize that only the intrinsic part is under consideration. The device is driven by four dc voltages, V_D , V_G , V_B , and V_S , defined with respect to some arbitrary reference point denoted by the ground symbol. Let four dc currents I_D , I_G , I_B , and I_S be defined as *entering* the device, as shown in the figure. We have seen that current flow is caused by the *transport* of electrons in the inversion layer (an *n*-channel device is assumed). Defining the *transport current* (or “conduction” current) as flowing from the drain through the channel to the source and denoting it by I_T , we have

$$I_D = I_T \quad (7.2.1a)$$

$$I_G = 0 \quad (7.2.1b)$$

$$I_B = 0 \quad (7.2.1c)$$

$$I_S = -I_T \quad (7.2.1d)$$

where we have neglected the extremely small leakage current through the insulator and the small leakage current through the depletion region. We have already derived

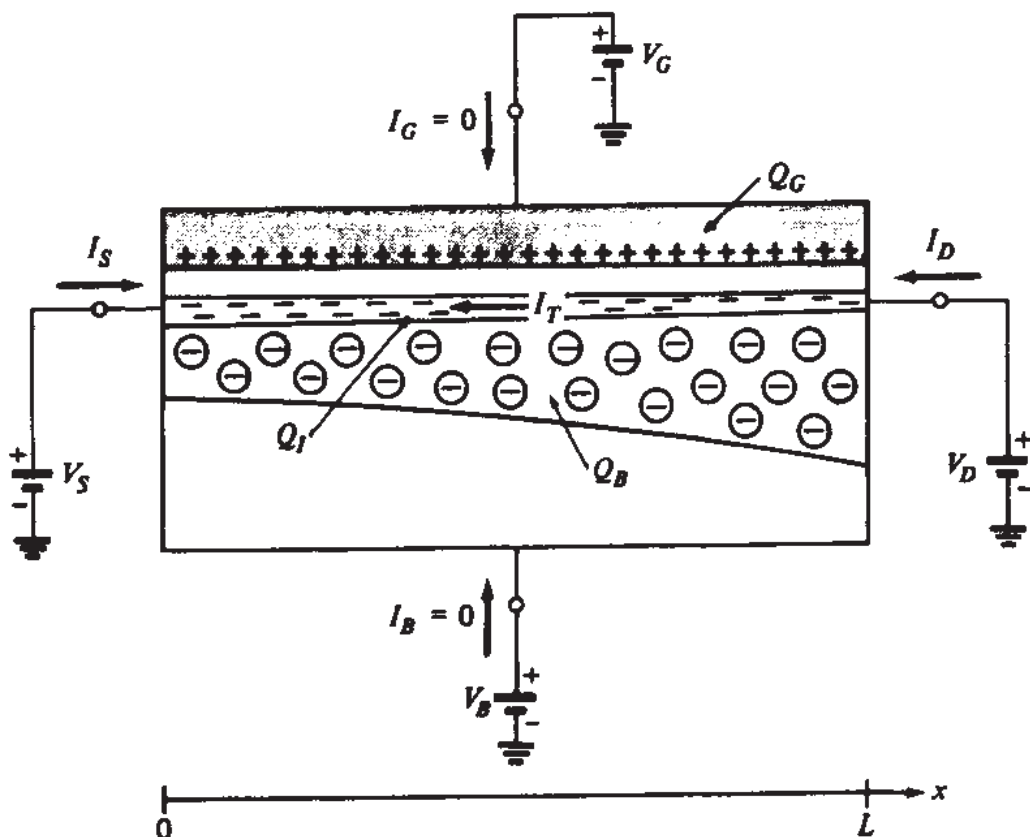


FIGURE 7.2
Definition of currents and charges under dc excitation.

several expressions for I_T (called I_{DS} in Chap. 4), which can be written in the general form†:

$$I_T = h_T(V_D, V_G, V_B, V_S) \quad (7.2.2)$$

where $h_T(\)$ is an appropriate function dependent on the model used to describe the device's dc behavior. Such models have been discussed in Chap. 4.

In previous chapters we have considered the inversion layer, gate, and depletion region charges per unit area (Q'_I , Q'_G , and Q'_B). These quantities depend, in general, on the position x along the channel. We will need to determine the corresponding *total* charges Q_I , Q_G , and Q_B . Consider Q_I as an example. A chunk of the inversion layer, with width W and length Δx at position x along the channel, will contain a charge $Q'_I(W \Delta x)$, where Q'_I depends on x . Thus, the total inversion layer charge will be

$$Q_I = \int_0^L Q'_I W dx$$

†In most cases in previous chapters we have given the drain current as a function of voltages of the form V_{KL} , where K and L can denote any two transistor terminals. We can write $V_{KL} = V_K - V_L$, in which case an expression in the form of (7.2.2) is obtained. In general cases where the drain current expression includes surface potential values of the source and drain ends of the channel (Sec. 4.3), we can, in principle, use for those potentials an approximate *explicit* expression in terms of $V_{SB} = V_S - V_B$, $V_{DB} = V_D - V_B$, and $V_{GB} = V_G - V_B$; again, the result will be an explicit expression for the drain current in the form of (7.2.2).

or

$$Q_I = W \int_0^L Q'_I dx \quad (7.2.3a)$$

Similarly,

$$Q_G = W \int_0^L Q'_G dx \quad (7.2.3b)$$

$$Q_B = W \int_0^L Q'_B dx \quad (7.2.3c)$$

We will undertake the evaluation of these integrals in Sec. 7.4. For now, we only note that, as might be expected, the final results will depend on the terminal voltages

$$Q_I = f_I(V_D, V_G, V_B, V_S) \quad (7.2.4a)$$

$$Q_G = f_G(V_D, V_G, V_B, V_S) \quad (7.2.4b)$$

$$Q_B = f_B(V_D, V_G, V_B, V_S) \quad (7.2.4c)$$

Q_G and Q_B can be interpreted as charges “stored” in the device; however, *the interpretation of Q_I requires more care*. Q_I is due to electrons in the inversion layer. These electrons are not really stored in the device. They enter through the source and eventually leave through the drain, being continuously replaced by new electrons entering through the source. Q_I is simply the total charge of the electrons that happen to be in the inversion layer at any given instant. This quantity is constant in Fig. 7.2 despite the fact that the “individual electrons” giving rise to it may be different at different instants.

We will now allow the terminal voltages to vary with time. Total time-varying quantities will be denoted by lower-case symbols with upper-case subscripts, as shown in Fig. 7.3. We will assume that *the variation of the terminal voltages is sufficiently slow, so that the device operates quasi-statically*,^{5,8,10,15,16,18–21,25, 26,28–39,41–53, 55–58,60,61} as defined below:

Quasi-static operation assumption: Let $v_D(t)$, $v_G(t)$, $v_B(t)$, and $v_S(t)$ be the varying terminal voltages; then at any position, the charges per unit area at any time t' are assumed identical to those that would be found if dc voltages were used instead: $V_D = v_D(t')$, $V_G = v_G(t')$, $V_B = v_B(t')$, and $V_S = v_S(t')$.

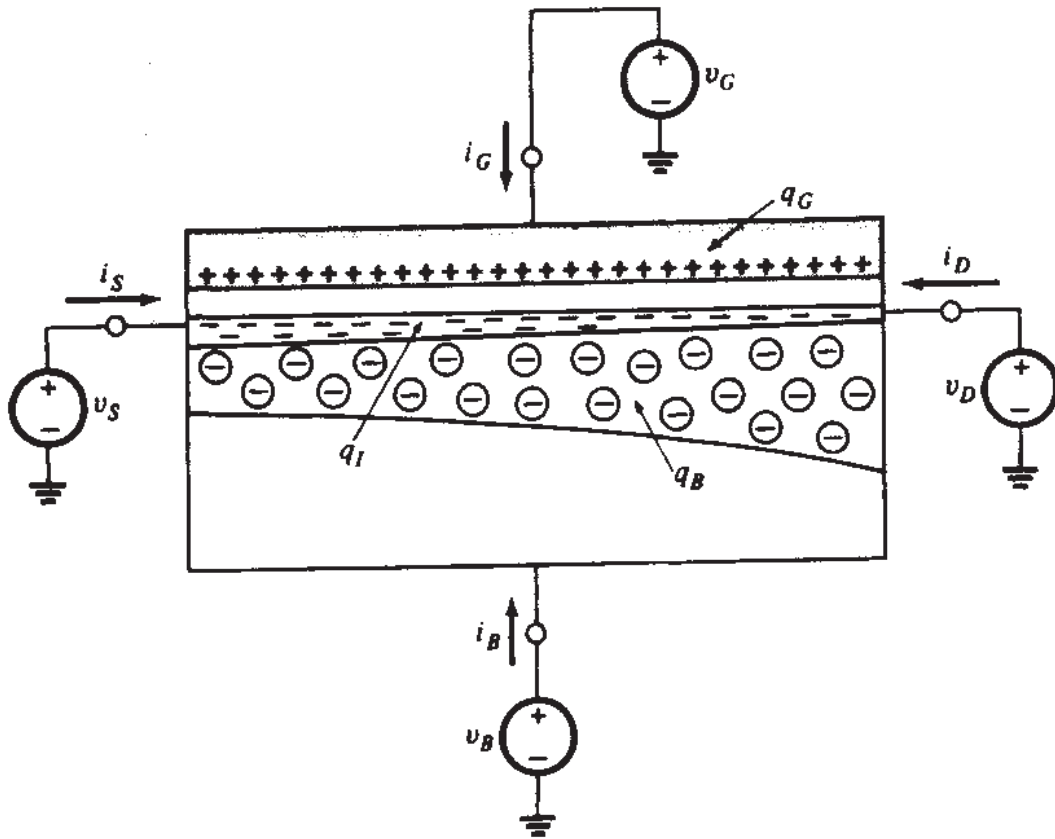


FIGURE 7.3

Definition of currents and charges in the presence of varying terminal voltages. Lowercase symbols with capital subscripts denote total time-varying quantities.

Under this assumption, the total charges q_I , q_G , and q_B can still be found from (7.2.3), and will be given by

$$q_I(t) = f_I(v_D(t), v_G(t), v_B(t), v_S(t)), \quad \text{quasi-static operation} \quad (7.2.5a)$$

$$q_G(t) = f_G(v_D(t), v_G(t), v_B(t), v_S(t)), \quad \text{quasi-static operation} \quad (7.2.5b)$$

$$q_B(t) = f_B(v_D(t), v_G(t), v_B(t), v_S(t)), \quad \text{quasi-static operation} \quad (7.2.5c)$$

where f_I , f_G , and f_B represent the *same* functions as in (7.2.4). However, in contrast to the calculation of the charges, the currents *cannot* be evaluated by using the current formulas for static operation. For example, since q_G is varying, there will be a nonzero gate current; this current cannot be predicted using the relation (7.2.1b); that equation is only valid for static operation. The proper way to calculate the currents in quasi-static operation will be discussed shortly.

It may be intuitively clear that the assumption of quasi-static operation will fail if the terminal voltages vary too fast. For example, if a step waveform is used for one of them, the charges will exhibit some inertia and cannot be expected to readjust themselves instantaneously. Such cases are not considered in this and the next several sections. Instead, we are assuming that the terminal voltages vary sufficiently slowly for the quasi-static approximation to be valid. The limits of validity of the quasi-static

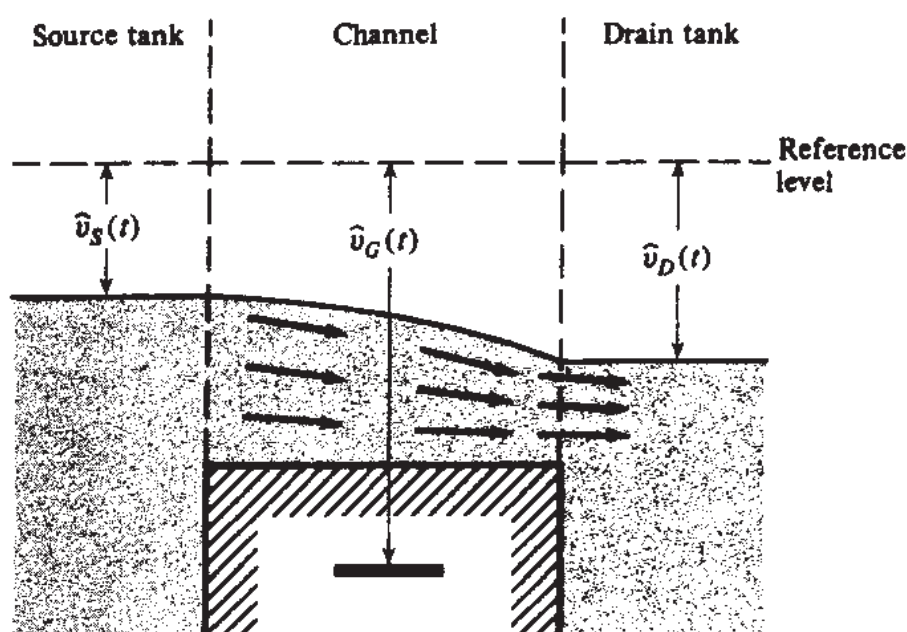


FIGURE 7.4

A fluid dynamical analog for illustrating charge motion in a transistor.

approximation, and the reasons for its failing in extreme cases, will be discussed in Sec. 7.6. Non-quasi-static analysis will be discussed in Sec. 7.7 and, for small signals, in Sec. 9.4.

One can obtain intuition about the quasi-static approximation by using the fluid dynamical analog introduced in Sec. 1.6. Consider the case shown in Fig. 7.4. Assume the piston is moving, and let $\hat{v}_G(t)$ be the depth of its handle below the reference level. If the piston is moving sufficiently slowly, the fluid distribution at any instant t' will be practically as if $\hat{v}_G(t)$ has been frozen permanently at the value $\hat{v}_G(t')$. Then the quasi-static approximation holds for the fluid dynamical analog. Obviously, if $\hat{v}_G(t)$ varies fast, this will no longer be the case, since the fluid is not given enough time to accommodate itself. Similar comments hold if \hat{v}_S or \hat{v}_D is varied.

7.3 TERMINAL CURRENTS IN QUASI-STATIC OPERATION

We will now evaluate the terminal currents of the *idealized* device in Fig. 7.3, assuming *quasi-static operation*. (The total terminal currents of a real device will include current components owing to extrinsic parasitic capacitances. Such capacitances are discussed in Sec. 8.4.)

Assuming there is no gate “leakage” and, therefore, no gate transport current, all gate current in Fig. 7.3 is associated with a changing gate charge:

$$i_G(t) = \frac{dq_G}{dt} \quad (7.3.1)$$

Similarly, assuming no leakage in the bulk, the transport current in the depletion region is zero. Then all bulk current is associated with free electrons changing the charge in that region, by depleting or “covering” acceptor atoms there. More specifically, assume that the bulk charge increases by an amount Δq_B . This means that free electrons of total charge $-\Delta q_B$ must have left the acceptors they were previously covering, and must have exited through the substrate terminal (we assume no electron-hole recombination). This is equivalent to an opposite charge, $+\Delta q_B$, *entering* the device through the substrate terminal. Thus, the movement of mobile charges through the substrate terminal can be related to the change in the charge of the depletion region, and we can write

$$i_B(t) = \frac{dq_B}{dt} \quad (7.3.2)$$

Finally, the sum of the drain and source currents represents the total current entering the channel, which will change the inversion layer charge:^{5,8,10,15,16,18-21}

$$i_D(t) + i_S(t) = \frac{dq_I}{dt} \quad (7.3.3)$$

We need expressions for $i_D(t)$ and $i_S(t)$ separately. *The reader is cautioned that, under the assumption of quasi-static operation, conflicting ways to evaluate these quantities are suggested in the literature.* This is because of conflicting interpretations of the “charges associated with the drain and with the source” quantities to be discussed below. We emphasize that just looking at equations can be very misleading in the development of relations for $i_D(t)$ and $i_S(t)$. *A correct interpretation of the charges appearing in such equations is essential.* This requires making several fine points as explained below.

We begin by noting that at dc, $i_D(t) = -i_S(t)$ [see (7.2.1)]. Hence, from (7.3.3), $dq_I/dt = 0$; that is, q_I is constant. In the general case of time-varying voltages, though, q_I will be varying too. Then $dq_I/dt \neq 0$, and $i_D(t)$ *cannot* be equal to $-i_S(t)$ in (7.3.3). To illustrate this point, assume that in Fig. 7.3 voltages v_D , v_S , and v_B are kept constant but v_G is varying as in Fig. 7.5a. We will assume that the variations of v_G are slow enough so that quasi-static operation is maintained. Then $q_I(t)$ will be given by an equation of the form (7.2.5a); its magnitude is plotted in Fig. 7.5b. The corresponding currents $i_D(t)$ and $-i_S(t)$ are shown in Fig. 7.5c. To interpret this figure, note that $i_D(t)$ is a measure of the number of *electrons leaving* the device through the drain per unit time and $-i_S(t)$ is a measure of the number of *electrons entering* the device through the source per unit time.† If “dc-like” behavior were obeyed at all times, we

† Recall that negative charges moving from left to right correspond to positive current flowing from right to left.

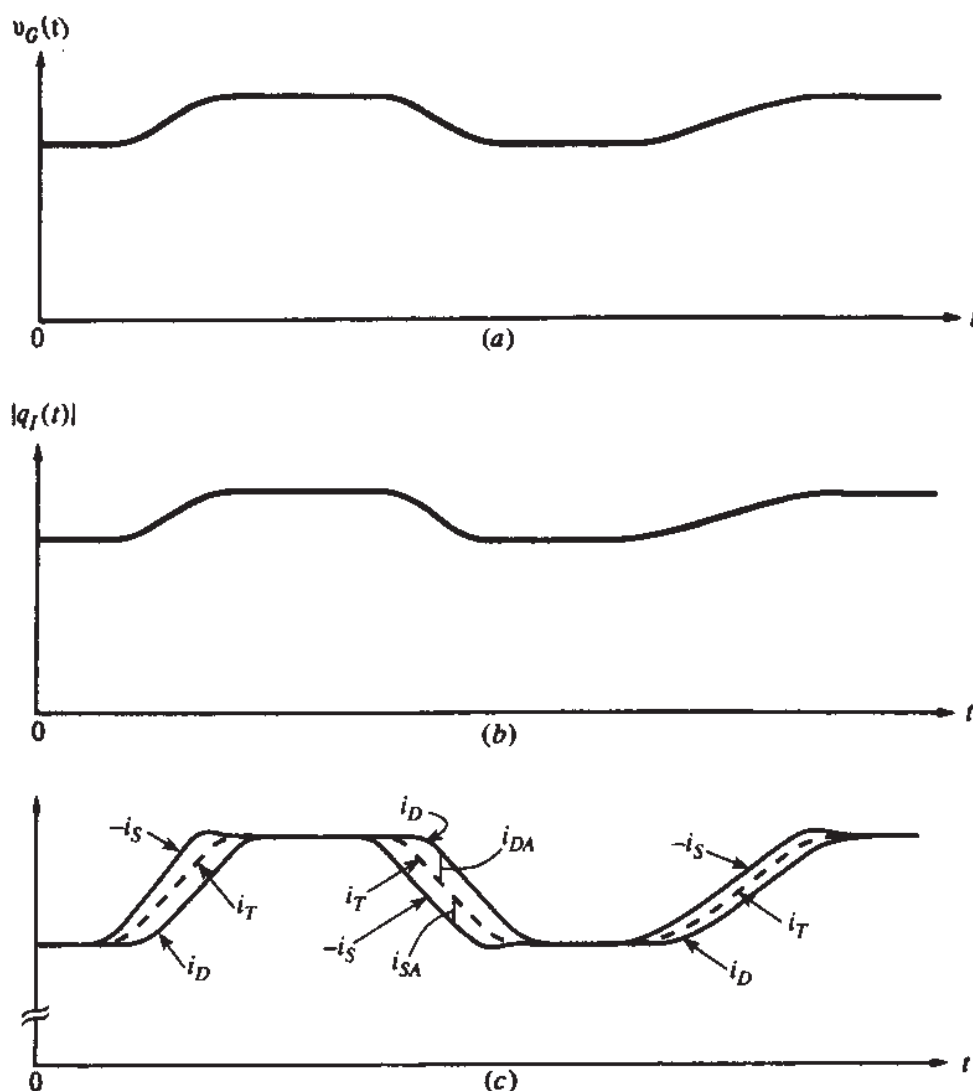


FIGURE 7.5

(a) Gate voltage waveform for the device of Fig. 7.3; all other terminal voltages are assumed fixed; (b) corresponding inversion layer charge magnitude waveform assuming quasi-static operation; (c) total drain current and the negative of the total source current; the transport and charging current components are also shown.

would have $-i_S(t) = i_D(t) = i_T(t)$. The quantity $i_T(t)$ would be given by (7.2.2) after replacing V_G in it by $v_G(t)$. This is shown by the dashed line in Fig. 7.5c. However, what we actually have is shown by the *solid* lines and can be explained as follows. When $v_G(t)$ is *increasing*, $|q_I(t)|$ must increase, as shown in Fig. 7.5b. Thus the number of electrons in the channel must increase. For this to happen, the rate of supply of electrons from the source, $-i_S(t)$, must temporarily become larger than the rate of removal of electrons from the drain, $i_D(t)$. As can be deduced from the quantitative results presented later, this is accomplished by $-i_S(t)$ becoming temporarily larger than $i_T(t)$, and by $i_D(t)$ becoming temporarily smaller than $i_T(t)$, as in Fig. 7.5c. These comments hold for both up-going transitions in the figure.

If $v_G(t)$ is *decreasing* instead, the opposite will be true. Here $|q_I(t)|$, and thus the number of electrons in the channel, must decrease. For this to happen, the rate of removal from the drain must temporarily exceed the rate of supply from the source. This is accomplished by $i_D(t)$ temporarily becoming larger than $i_T(t)$, and by $-i_S(t)$ temporarily becoming smaller than $i_T(t)$, as shown in the falling part of the plots in Fig. 7.5c.

The difference between the curve for i_D and the curve for $-i_S$ in Fig. 7.5c must be equal to the total rate of change of q_I , as follows from (7.3.3). By integrating (7.3.3) it can be deduced that the area between the solid lines at each transition is equal to the total change in $|q_I|$. If $|q_I(t)|$ in Fig. 7.5b goes up, down, and up again by equal amounts, the three areas enclosed by solid lines at each transition in Fig. 7.5c will all be equal. However, because the first up-going transition is faster than the second up-going transition, i_D and $-i_S$ need to deviate more from i_T to accomplish the same change in q_I .

From the above arguments it follows that, in general, $i_D(t)$ and $i_S(t)$ will differ from their transport values $i_T(t)$ and $-i_T(t)$, respectively. Denoting the differences by $i_{DA}(t)$ and $i_{SA}(t)$, we can write

$$\boxed{i_D(t) = i_T(t) + i_{DA}(t)} \quad (7.3.4a)$$

$$\boxed{i_S(t) = -i_T(t) + i_{SA}(t)} \quad (7.3.4b)$$

From these equations and (7.3.3) we have

$$i_{DA}(t) + i_{SA}(t) = \frac{dq_I}{dt} \quad (7.3.5)$$

Thus, we can conveniently view $i_T(t)$ as wholly responsible for the transport effect, and $i_{DA}(t)$, $i_{SA}(t)$ as wholly responsible for changing q_I . (For this reason these two currents are sometimes called “charging” currents.) Of course, it should be clear that we cannot identify particular electrons as causing transport current and others as charging the inversion layer. All electrons entering the source can be envisioned to eventually leave through the drain [see comment following (7.2.4c)]. However, since $i_D(t) \neq -i_S(t)$, the behavior an external observer sees is indistinguishable from what one *would* have if some of the charge went to contribute to a transport current while the rest went to change q_I , the latter often viewed as “stored” in the inversion layer at any given instant. To make this imaginary picture complete, let us associate $i_{DA}(t)$ and $i_{SA}(t)$ with two *fictitious* charges. If $i_{DA}(t)$ causes the inversion layer charge to change by an amount Δq_D in time Δt , we can write

$$\boxed{i_{DA}(t) = \frac{dq_D}{dt}} \quad (7.3.6a)$$

and, if $i_{SA}(t)$ causes the inversion layer charge to change by an amount Δq_S in time Δt , we can write†

$$i_{SA}(t) = \frac{dq_S}{dt} \quad (7.3.6b)$$

Expressions for q_D and q_S will be given shortly. The last three equations give

$$\frac{dq_D}{dt} + \frac{dq_S}{dt} = \frac{dq_I}{dt} \quad (7.3.7)$$

Intuition about the above ideas can be increased with the help of the fluid dynamical analog in Fig. 7.4. Assume \hat{v}_G has been fixed for a long time, so that the fluid has attained a steady-state distribution and the total rate of flow into the channel from the source tank is equal to the rate of flow out of the channel and into the drain tank. Now assume that \hat{v}_G is slowly increased (the piston is slowly moved downward). The total amount of fluid in the channel must also increase, which, of course, cannot happen if the two rates of flow continue to be equal. The rate of flow from the source tank will thus temporarily become larger than what would be predicted from “dc considerations,”‡ while the rate of exit from the channel and into the drain tank will temporarily become smaller than what would be predicted from “dc considerations.” (Note by the way that, if the movement of the piston downward is sufficiently slow, then at all times the flow of water is from left to right. Although the rate of flow is smaller at the drain end, no water is seen to flow backward from the drain tank into the channel.) If the piston is slowly raised (\hat{v}_G is decreasing), the total amount of fluid in the channel must decrease. The rate of flow from the source tank will become smaller while the rate of exit into the drain tank will become larger than the rate that would be predicted by “dc considerations.”

Note that dq_S/dt and dq_D/dt in the above equations correspond only to the *differences* between the actual and the dc-like values of flow at the source and the drain; they do *not* represent total flow. For example, $dq_D/dt < 0$ should not be interpreted to imply that the drain current is negative. It only implies that the actual drain current $i_D(t)$ is less than the corresponding transport value $i_T(t)$, as can be deduced by using (7.3.6a) in (7.3.4a).

†The definitions of q_D and q_S here are purposely imprecise. Until further notice, q_D and q_S can be any functions the time derivatives of which predict the correct values for $i_{DA}(t)$ and $i_{SA}(t)$, respectively. The reader is urged to avoid assigning physical significance to q_D and q_S at this point. A discussion of this will follow.

‡The “rate that would be predicted from dc considerations” at time $t = t'$ is what the rate would have been if $\hat{v}_G(t)$ had been frozen for a long time at the value $\hat{v}_G(t')$.

To determine $i_D(t)$ and $i_S(t)$ in (7.3.4) and (7.3.6) we need dq_D/dt and dq_S/dt . Any functions $q_D(t)$ and $q_S(t)$ giving correct time derivatives will do for this purpose. An obvious simple choice is to define these functions so that¹⁶

$$q_D(t) + q_S(t) = q_I(t) \quad (7.3.8)$$

Because of this choice, the following interpretation is sometimes tempting, albeit *not* accurate: "The total charge stored in the inversion layer consists of two components, one that has come through the drain, and one that has come through the source." The reasons that this interpretation is not accurate are as follows. First, the view of q_I as stored charge leaves much to be desired, as explained in the paragraph following (7.2.4c). Second, as can be deduced from the comment preceding (7.3.8), q_D and q_S are not unique, so it is not justified to assign to them a unique physical significance. Third, in the general case it is not correct to identify q_D and q_S as necessarily "coming" from the drain and source respectively. For example, in the case illustrated in Fig. 7.5 in conjunction with Fig. 7.3, and provided the changes in $v_G(t)$ are sufficiently slow, electrons travel from the source through the channel to the drain at all times (just like the water in Fig. 7.4). In this picture, no electrons travel backward from the drain into the channel. Thus, all charges "come from the source," and no charge "comes from the drain." That does not say anything about the value of q_D , which can still be nonzero and is such that its derivative, from (7.3.6a) and (7.3.4a), is equal to $i_D(t) - i_T(t)$, this difference being indicated by i_{DA} in Fig. 7.5c. It is thus better to think of q_D as the integral of this difference and not to assign a further physical significance to it. Similar comments apply to q_S . In fact, a complete development is possible dealing with i_{DA} and i_{SA} directly, without ever defining q_D and q_S . However, in conformance with the literature we will continue using these quantities, and we will assume that they are defined in such a way that (7.3.8) holds.†

Various approaches have been used in the literature for evaluating q_D and q_S , or i_{DA} and i_{SA} .^{15,16,18-21,24-30,34-37,44,70} Here we will adopt an approach that can be rigorously shown to be correct²¹ and which has been demonstrated to agree with experiment. (More on the correctness of the approach will follow later.) We begin by assuming dc operation and define two charges Q_D and Q_S as follows:

$$Q_D = W \int_0^L \frac{x}{L} Q'_I dx \quad (7.3.9a)$$

$$Q_S = W \int_0^L \left(1 - \frac{x}{L}\right) Q'_I dx \quad (7.3.9b)$$

where Q'_I is the inversion layer charge per unit area. The sum of Q_D and Q_S is seen to be equal to the total inversion layer charge Q_I [which is given by (7.2.3a)]. The evalu-

†A possible definition of q_D and q_S is discussed in Appendix L.

ation of the above integrals will be considered in Sec. 7.4. For now we only mention that the results will be explicit functions of the terminal voltages:

$$Q_D = f_D(V_D, V_G, V_B, V_S) \quad (7.3.10a)$$

$$Q_S = f_S(V_D, V_G, V_B, V_S) \quad (7.3.10b)$$

If the terminal voltages are allowed to vary, one can evaluate the right-hand sides in (7.3.9) by using $q'_i(t)$ in lieu of Q'_i . If the variation of the terminal voltages is slow enough so that *quasi-static* operation is maintained, the results of this evaluation will be two quantities, denoted by $q_D(t)$ and $q_S(t)$, given by

$$q_D(t) = f_D(v_D(t), v_G(t), v_B(t), v_S(t)), \quad \text{quasi-static operation} \quad (7.3.11a)$$

$$q_S(t) = f_S(v_D(t), v_G(t), v_B(t), v_S(t)), \quad \text{quasi-static operation} \quad (7.3.11b)$$

where f_D and f_S represent the *same* functions as in (7.3.10). This is consistent with the definition of quasi-static operation given in Sec. 7.2. Thus, $q_D(t)$ and $q_S(t)$ as defined here satisfy (7.3.8). It can now be shown²¹ (Appendix L) that, for the long-channel devices we are considering, the instantaneous currents $i_D(t)$ and $i_S(t)$ in quasi-static operation will be given by (7.3.4) and (7.3.6), with $q_D(t)$ and $q_S(t)$ as given by (7.3.11), and

$$i_T(t) = h_T(v_D(t), v_G(t), v_B(t), v_S(t)) \quad (7.3.12)$$

where h_T represents the *same* function as in (7.2.2). The proof relies on the “continuity equation,” which is introduced in Sec. 7.7, and carefully considers the current and charge at each point in the channel. The evidence that the above approach is correct is overwhelming:

1. It is physically and mathematically sound. This fact can be appreciated if the quantities i_{DA} , i_{SA} , q_D , and q_S are carefully interpreted as explained earlier, and the detailed development²¹ is followed (Appendix L).
2. It agrees with experiment,²¹ as long as the speed of variation is maintained sufficiently low for the quasi-static operation assumption to be valid.
3. Other approaches using perturbation techniques²⁵ give equivalent results.³⁰
4. Non-quasi-static large-scale numerical models, applied in the special case of quasi-static operation, give equivalent results.²⁰
5. Small-signal models derived by using the above approach agree with non-quasi-static small-signal models in the special case of quasi-static operation (Chap. 9).

It should be noted that models based on the assumption of quasi-static operation are sometimes abused: they are used to predict currents in very high speed operation, in which the device does not operate quasi-statically. In such cases, the above theory is found not to provide satisfactory results (it should not be expected to, either).

As follows from the above, the transport component can be found by using the models of Chap. 4 and will not be considered further here. To simplify our discus-

sion, we will consider only the “charging” current components. We thus concentrate on the following equations:

$$i_{DA}(t) = \frac{dq_D}{dt} \quad (7.3.13a)$$

$$i_G(t) = \frac{dq_G}{dt} \quad (7.3.13b)$$

$$i_B(t) = \frac{dq_B}{dt} \quad (7.3.13c)$$

$$i_{SA}(t) = \frac{dq_S}{dt} \quad (7.3.13d)$$

It is interesting to note at this point that not only will Kirchhoff's current law hold for the total currents,

$$i_D(t) + i_G(t) + i_B(t) + i_S(t) = 0 \quad (7.3.14)$$

but it will also hold for the charging currents as is obvious by using (7.3.4) in the above equation:

$$i_{DA}(t) + i_G(t) + i_B(t) + i_{SA}(t) = 0 \quad (7.3.15)$$

Using (7.2.5b), (7.2.5c), and (7.3.11) in (7.3.13), and applying the chain rule of differentiation, we obtain

$$i_{DA}(t) = \frac{\partial q_D}{\partial v_D} \frac{dv_D}{dt} + \frac{\partial q_D}{\partial v_G} \frac{dv_G}{dt} + \frac{\partial q_D}{\partial v_B} \frac{dv_B}{dt} + \frac{\partial q_D}{\partial v_S} \frac{dv_S}{dt} \quad (7.3.16a)$$

$$i_G(t) = \frac{\partial q_G}{\partial v_D} \frac{dv_D}{dt} + \frac{\partial q_G}{\partial v_G} \frac{dv_G}{dt} + \frac{\partial q_G}{\partial v_B} \frac{dv_B}{dt} + \frac{\partial q_G}{\partial v_S} \frac{dv_S}{dt} \quad (7.3.16b)$$

$$i_B(t) = \frac{\partial q_B}{\partial v_D} \frac{dv_D}{dt} + \frac{\partial q_B}{\partial v_G} \frac{dv_G}{dt} + \frac{\partial q_B}{\partial v_B} \frac{dv_B}{dt} + \frac{\partial q_B}{\partial v_S} \frac{dv_S}{dt} \quad (7.3.16c)$$

$$i_{SA}(t) = \frac{\partial q_S}{\partial v_D} \frac{dv_D}{dt} + \frac{\partial q_S}{\partial v_G} \frac{dv_G}{dt} + \frac{\partial q_S}{\partial v_B} \frac{dv_B}{dt} + \frac{\partial q_S}{\partial v_S} \frac{dv_S}{dt} \quad (7.3.16d)$$

To evaluate the above currents, we need expressions for the charges as functions of the terminal voltages.† Such expressions are developed in the following section.

†The implementation of (7.3.16) in a computer simulator requires care, in order to avoid results inconsistent with charge conservation.^{27,34-37,44,70}

7.4 EVALUATION OF CHARGES IN QUASI-STATIC OPERATION

7.4.1 Introduction

The charge expressions needed to complete the evaluation of the charging currents in (7.3.16) can be developed in a straightforward manner under the assumption of *quasi-static operation* (Sec. 7.2). Thus, we observe that, under this assumption, the expressions for $q_G(t)$, $q_B(t)$, $q_I(t)$, $q_D(t)$, and $q_S(t)$ in terms of $v_G(t)$, $v_B(t)$, $v_S(t)$, and $v_D(t)$ are identical to the expressions for the charges under dc conditions Q_G , Q_B , Q_I , Q_D , and Q_S in terms of the dc voltages V_G , V_B , V_S , and V_D . The latter expressions are given in the form of integrals in (7.2.3) and (7.3.9). If each region of operation is considered separately, these integrals lead to simple functions of the terminal voltages, with the exception of the moderate-inversion region. This is shown in the next several subsections. A more general evaluation is also possible, corresponding to the general charge sheet models of Sec. 4.3. This is discussed in Sec. 7.4.5.

The emphasis in this chapter is on illustrating the principles involved; in order to keep the expressions simple, we assume no short-channel effects are present. Using the same principles, such effects can be incorporated at the expense of complexity.^{49,52,53,56}

7.4.2 Strong Inversion

GENERAL EXPRESSIONS FOR NONSATURATION. The integrations indicated in (7.2.3) and (7.3.9) cannot be carried out directly since we do not have the corresponding charges per unit area as functions of x . Accordingly, we first perform a change of variables from the position x to V_{CB} , the strong-inversion “effective reverse bias” of the inversion layer at point x , with respect to the substrate. From (4.5.6) we have

$$dx = -\frac{\mu W}{I_{DSN}} Q'_I dV_{CB} \quad (7.4.1)$$

where I_{DSN} is the drain current in the nonsaturation region. Using the above relation, the variable of integration is changed in (7.2.3b) to give

$$\begin{aligned} Q_G &= W \int_{V_{SB}}^{V_{DB}} Q'_G \left(-\frac{\mu W}{I_{DSN}} Q'_I \right) dV_{CB} \\ &= -\frac{\mu W^2}{I_{DSN}} \int_{V_{SB}}^{V_{DB}} Q'_G Q'_I dV_{CB} \end{aligned} \quad (7.4.2a)$$

Similarly, (7.2.3c) becomes:

$$Q_B = -\frac{\mu W^2}{I_{DSN}} \int_{V_{SB}}^{V_{DB}} Q'_B Q'_I dV_{CB} \quad (7.4.2b)$$

and (7.2.3a) becomes

$$Q_I = -\frac{\mu W^2}{I_{DSN}} \int_{V_{SB}}^{V_{DB}} Q_I'^2 dV_{CB} \quad (7.4.2c)$$

Q_D and Q_S from (7.3.9) become

$$Q_D = -\frac{\mu W^2}{I_{DSN}} \int_{V_{SB}}^{V_{DB}} \frac{x}{L} Q_I'^2 dV_{CB} \quad (7.4.3a)$$

$$Q_S = -\frac{\mu W^2}{I_{DSN}} \int_{V_{SB}}^{V_{DB}} \left(1 - \frac{x}{L}\right) Q_I'^2 dV_{CB} \quad (7.4.3b)$$

Finally, to express x in the above equations in terms of V_{CB} we integrate (7.4.1) from $x = 0$ to an arbitrary point in the channel:

$$x = -\frac{\mu W}{I_{DSN}} \int_{V_{SB}}^{V_{CB}} Q_I' dU_{CB} \quad (7.4.4)$$

where U_{CB} is a dummy variable of integration. A result equivalent to the above has been obtained in Sec. 4.5.3. Thus from (4.5.48) we have

$$x = L \frac{h(V_{GB}, V_{SB}, V_{CB})}{h(V_{GB}, V_{SB}, V_{DB})} \quad (7.4.5)$$

where h represents the function multiplying W/L in the nonsaturation drain current expression:

$$I_{DSN} = \frac{W}{L} h(V_{GB}, V_{SB}, V_{DB}) \quad (7.4.6)$$

It is easy to show that (7.4.5) is equivalent to (7.4.4) (Prob. 7.1).

GENERAL EXPRESSIONS INCLUDING SATURATION. Let the value of V_{DB} at which the transistor enters saturation be denoted by V_P . Since the channel is assumed long, if V_{DB} is raised above V_P , the conditions in the channel will remain practically unaffected (Sec. 4.5.1). Thus, for example, let $g_I(V_{GB}, V_{SB}, V_{DB})$ represent the expression giving Q_I in nonsaturation, as it results from performing the integration in (7.4.2c). We will have

$$Q_I = \begin{cases} g_I(V_{GB}, V_{SB}, V_{DB}), & V_{DB} \leq V_P \\ g_I(V_{GB}, V_{SB}, V_P), & V_{DB} > V_P \end{cases} \quad (7.4.7a)$$

$$(7.4.7b)$$

Corresponding relations can be written for Q_G , Q_B , Q_D , and Q_S .

Using the above results, expressions of varying complexity can be found for Q_G , Q_B , Q_D , and Q_S depending on the complexity of the model used for Q'_G , Q'_B , Q'_D , Q'_S , and I_{DSN} .

SIMPLIFIED MODEL. The computationally efficient simplified model derived in Sec. 4.5.3 resulted in expression (4.5.39) for the drain-to-source current:

$$I_{DS} = I'_{DS}(1 - \eta^2) \quad (7.4.8)$$

where

$$I'_{DS} = \frac{W}{L} \mu C'_{ox} \frac{(V_{GS} - V_T)^2}{2\alpha} \quad (7.4.9)$$

and η was a parameter defined in a way that makes (7.4.8) valid in *both* nonsaturation and saturation:

$$\eta = \begin{cases} 1 - \frac{V_{DS}}{V'_{DS}}, & V_{DS} \leq V'_{DS} \\ 0, & V_{DS} > V'_{DS} \end{cases} \quad (7.4.10a)$$

$$(7.4.10b)$$

where

$$V'_{DS} = \frac{V_{GS} - V_T}{\alpha} \quad (7.4.11)$$

The parameter η has been plotted in Fig. 4.20.

To develop the above model, the following expressions were used in Sec. 4.5.3 for the inversion layer and depletion region charges per unit area, in nonsaturation:

$$Q'_I = -C'_{ox} [V_{GB} - V_{SB} - V_T - \alpha(V_{CB} - V_{SB})] \quad (7.4.12)$$

$$Q'_B = -C'_{ox} [\gamma\sqrt{\phi_0 + V_{SB}} + (\alpha - 1)(V_{CB} - V_{SB})] \quad (7.4.13)$$

We can thus find the corresponding nonsaturation *total* charges by using these in (7.4.2c) and (7.4.2b).²⁹ We can put the results in a form valid in *both* nonsaturation and saturation by using the convenient parameter η defined in (7.4.10). After some algebra, we obtain

$$Q_I = -WLC'_{ox}(V_{GS} - V_T) \frac{2}{3} \frac{1 + \eta + \eta^2}{1 + \eta} \quad (7.4.14)$$

$$Q_B = -WLC'_{ox} \left[\gamma\sqrt{\phi_0 + V_{SB}} + \frac{\alpha - 1}{\alpha} (V_{GS} - V_T) \left(1 - \frac{2}{3} \frac{1 + \eta + \eta^2}{1 + \eta} \right) \right] \quad (7.4.15)$$

The total gate charge can be found from (7.4.2a). However, since Q_I and Q_B have already been found, it is simpler to find Q_G by using the above results in the charge neutrality equation:

$$Q_G + Q_o + Q_I + Q_B = 0 \quad (7.4.16)$$

where Q_o is the total equivalent interface charge. The result is

$$Q_G = WLC'_{ox} \left[\frac{V_{GS} - V_T}{\alpha} \left(\alpha - 1 + \frac{2}{3} \frac{1 + \eta + \eta^2}{1 + \eta} \right) + \gamma \sqrt{\phi_0 + V_{SB}} \right] - Q_o \quad (7.4.17)$$

To find Q_D and Q_S from (7.4.3) we need to relate x to V_{CB} . This can be done by using (7.4.4) or (7.4.5):

$$x = L \frac{(V_{GS} - V_T)(V_{CB} - V_{SB}) - \frac{1}{2} \alpha (V_{CB} - V_{SB})^2}{(V_{GS} - V_T)(V_{DB} - V_{SB}) - \frac{1}{2} \alpha (V_{DB} - V_{SB})^2} \quad (7.4.18)$$

Using this and (7.4.12) in (7.4.3a), and extending the result to cover the saturation region, we obtain

$$Q_D = -WLC'_{ox}(V_{GS} - V_T) \frac{4 + 8\eta + 12\eta^2 + 6\eta^3}{15(1 + \eta)^2} \quad (7.4.19)$$

Similarly, Q_S can be found by using (7.4.3b). However, it is simpler here to obtain it from $Q_S + Q_D = Q_I$, using (7.4.14) and (7.4.19). The result is

$$Q_S = -WLC'_{ox}(V_{GS} - V_T) \frac{6 + 12\eta + 8\eta^2 + 4\eta^3}{15(1 + \eta)^2} \quad (7.4.20)$$

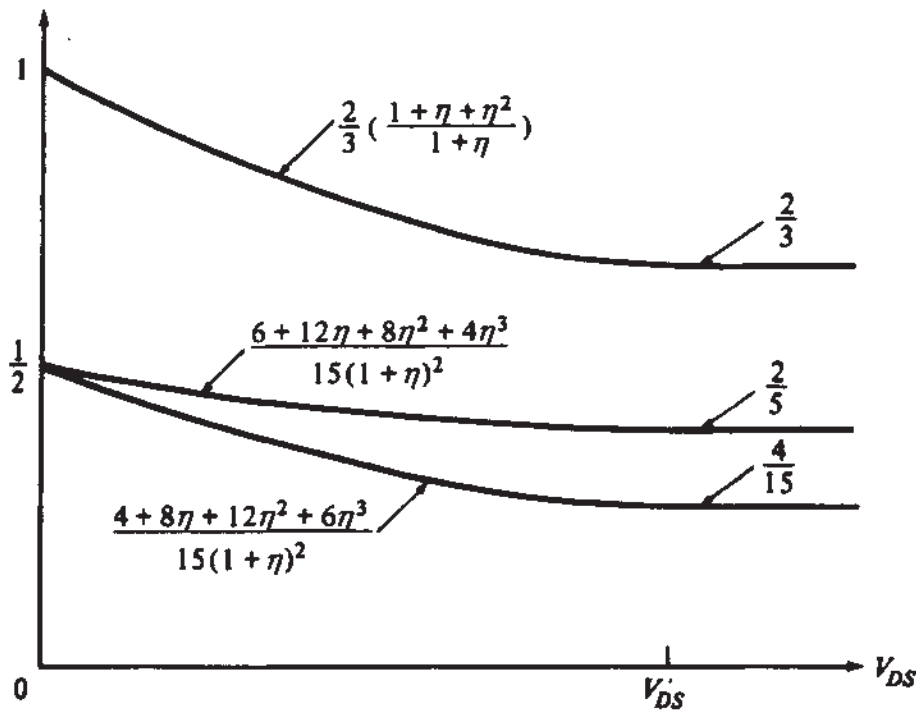
Plots of certain quantities appearing in (7.4.14), (7.4.15), (7.4.19), and (7.4.20) are shown in Fig. 7.6. As seen, despite the rather complicated look of the expressions, the form of the plots is rather simple. The reader may want to develop simpler functions of η which would approximate these plots with good accuracy.

Plots of the total charges as a function of V_{DS} , with V_{GS} as a parameter, are shown in Fig. 7.7.

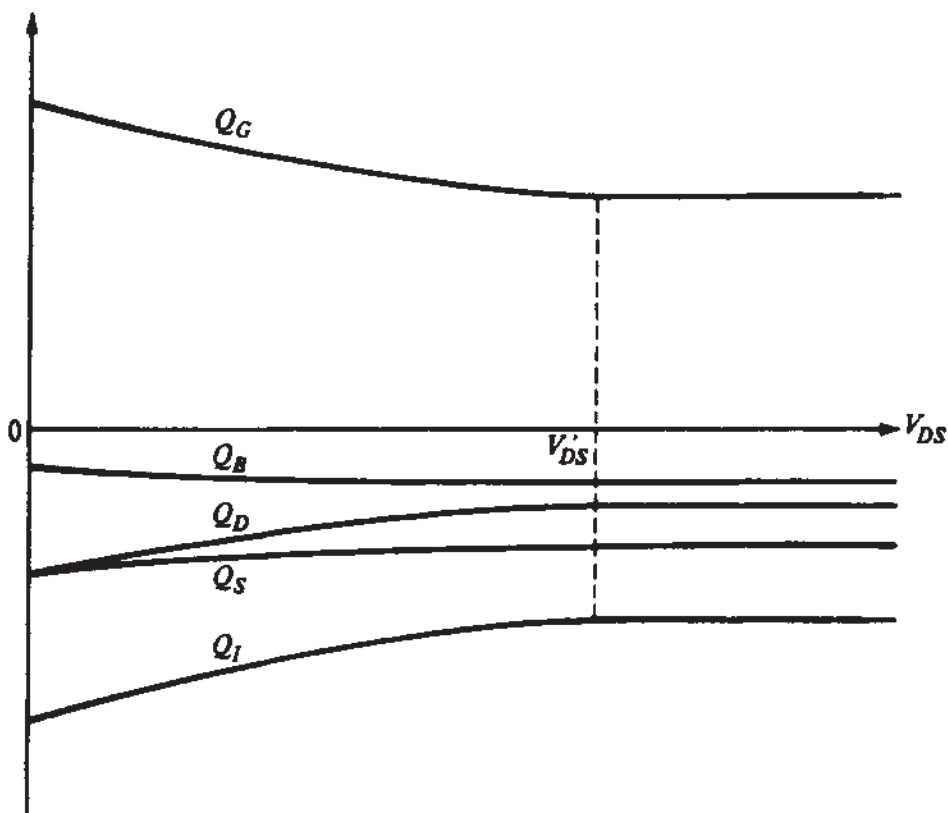
As a check of the above calculations, let us determine the charges at $V_{DS} = 0$ ($\eta = 1$):

$$Q_B|_{V_{DS}=0} = -WLC'_{ox} \gamma \sqrt{\phi_0 + V_{SB}} \quad (7.4.21)$$

$$Q_I|_{V_{DS}=0} = -WLC'_{ox}(V_{GS} - V_T) \quad (7.4.22)$$

**FIGURE 7.6**

Quantities used in total charge expressions versus drain-source voltage for fixed values of V_{GS} and V_{SB} .

**FIGURE 7.7**

Total charges versus drain-source voltage for fixed values of V_{GS} and V_{SB} .

$$Q_D|_{V_{DS}=0} = -\frac{WLC'_{ox}(V_{GS} - V_T)}{2} \quad (7.4.23)$$

$$Q_S|_{V_{DS}=0} = -\frac{WLC'_{ox}(V_{GS} - V_T)}{2} \quad (7.4.24)$$

$$Q_G|_{V_{DS}=0} = WLC'_{ox}[(V_{GS} - V_T) + \gamma\sqrt{\phi_0 + V_{SB}}] - Q_o \quad (7.4.25)$$

These equations make sense. Since $V_{DS} = 0$, the depletion region charge per unit area is uniform and is given by (7.4.13) with $V_{CB} = V_{SB}$. Multiplying this by the channel area WL gives (7.4.21). Similarly, (7.4.22) is simply the channel area times the uniform inversion layer charge per unit area as given from (7.4.12), with $V_{CB} = V_{SB}$. Q_D and Q_S are seen to be half of Q_I each, which makes sense because of symmetry. Finally, (7.4.25) can result from (7.4.16), (7.4.21), and (7.4.22).

In the saturation region ($\eta = 0$) we obtain

$$Q_{B,sat} = -WLC'_{ox}\left[\gamma\sqrt{\phi_0 + V_{SB}} + \frac{\alpha - 1}{3\alpha}(V_{GS} - V_T)\right] \quad (7.4.26)$$

$$Q_{I,sat} = -\frac{2}{3}WLC'_{ox}(V_{GS} - V_T) \quad (7.4.27)$$

$$Q_{D,sat} = -\frac{4}{15}WLC'_{ox}(V_{GS} - V_T) \quad (7.4.28)$$

$$Q_{S,sat} = -\frac{2}{5}WLC'_{ox}(V_{GS} - V_T) \quad (7.4.29)$$

$$Q_{G,sat} = WLC'_{ox}\left[\frac{V_{GS} - V_T}{\alpha}\left(\alpha - \frac{1}{3}\right) + \gamma\sqrt{\phi_0 + V_{SB}}\right] - Q_o \quad (7.4.30)$$

As seen, none of the above charges depends on the drain voltage. This is a manifestation of the fact that in saturation the drain can no longer have any influence on the intrinsic part of the device owing to pinchoff (excluding, of course, short-channel effects).

The influence of one terminal on the charge associated with another is in general nonreciprocal. For example, assume that V_S , V_B , and V_G are fixed and that V_D is varying in the saturation region. Since (7.4.30) is independent of V_D , the gate charge will remain fixed and no transient gate current will be observed. Now, assume that instead V_S , V_B , and V_D are fixed and that V_G is varying, again in the saturation region. From (7.4.28), it is apparent that Q_D will vary. Thus, a nonzero "charging current" will flow through the drain terminal in addition to the conduction current. These facts are apparent from (7.3.16a), (7.3.16b), (7.4.28), and (7.4.30). Such nonreciprocal influence of one terminal on another, although most strongly pronounced in saturation,

is also apparent in nonsaturation; it vanishes only at $V_{DS} = 0$. These effects will be considered in more detail in Chap. 9.

We note here that sometimes Q_D is assumed to be zero in the saturation region. This is justified by saying that in this region the channel is isolated from the drain. However, this isolation is only responsible for maintaining i_G independent of $v_D(t)$, as already argued. It is a fact that when $v_G(t)$ changes, $i_D(t)$ will also change, and there is no reason to assume that this change will not include a charging component $i_{DA}(t)$ (Sec. 7.3). In fact, measurements and numerical simulations show that, indeed, $i_{DA}(t)$ can be nonzero in saturation. From (7.3.6a) or (7.3.16a), then, q_D must be such as to give the correct value of $i_{DA}(t)$, which cannot happen if it is identically set to zero. So Q_D is *not* zero in saturation, although it is independent of v_D . There is *no* contradiction between “isolation” due to pinchoff and a nonzero value for Q_D , if Q_D is interpreted as explained in Sec. 7.3.

The expressions we have given so far for the charges are adequate for most quasi-static transient response calculations in strong inversion.

SYMMETRIC MODEL. The general equations (7.4.2) and (7.4.3) can be used to develop charge expressions corresponding to the complete symmetric model of Sec. 4.5.1. One must work in this case with the charges per unit area used in developing that model, which are given by (4.5.9) and (4.5.10). The resulting charge expressions have the attractive feature that they are completely symmetric with respect to V_{SB} and V_{DB} , just as was the case for the corresponding model for the current in Sec. 4.5.1. However, these expressions are rather involved. In particular, the expressions for Q_D and Q_S are very complicated,²¹ and one must resort to approximations. Since the principles have already been demonstrated, rather than hit the reader with more long formulas, we have summarized the charges for the accurate model in Appendix M.

7.4.3 Moderate Inversion

As was the case for drain current modeling, no simple closed-form expressions have been derived for the charges in the moderate-inversion region in terms of the terminal voltages. Some charge models neglect this region and assume that the weak-inversion and strong-inversion expressions hold in adjacent regions of V_{GS} . The limit point between these regions is often taken to be $V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F + V_{SB}}$, which is what we have denoted by V_M (the bottom of the moderate-inversion region). The resulting error in the moderate-inversion charges is not large. However, if it is attempted to differentiate these charges in order to obtain capacitance expressions (Chap. 8), large errors will result as we will see.

To model better the charges in moderate inversion, sometimes empirical expressions are used, resulting in curves which connect smoothly the weak inversion curves with those for strong inversion. When this is done, it is important to maintain continuity not only of the charges at the transition points but also of the *derivatives* of those charges with respect to the terminal voltages. In this way, the charge expressions will result in continuous capacitance expressions (Chap. 8). As is true with all

empirical expressions, it is also important that the parameters used in them be related to process parameters in a correct manner, so that predictions can be made in cases where experimental data are not available.

As will be seen in Sec. 7.4.5, it is possible to derive a model valid in all regions, including moderate inversion as a special case. However, as was the case with general drain current modeling, the price to be paid is computational complexity.

7.4.4 Weak Inversion

In weak inversion, the calculation of the charges is easy. First, we note that the depletion region charge per unit area is given by (4.3.14), repeated here:

$$Q'_B = -\gamma C'_{ox} \sqrt{\psi_s} \quad (7.4.31)$$

As seen in Sec. 4.6, in the weak-inversion region (as well as in depletion), the surface potential ψ_s is practically independent of position, and is given by

$$\psi_s \approx \psi_{sa} = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right)^2 \quad (7.4.32)$$

Thus, Q'_B is also independent of position. The total depletion charge is thus simply WLQ'_B , which gives

$$Q_B = -WLC'_{ox}\gamma \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right) \quad (7.4.33)$$

Consider now the charge neutrality equation $Q_G + Q_o + Q_I + Q_B = 0$. For the purposes of calculating Q_G , we can use the fact that in weak inversion $Q_I \ll Q_B$ (Sec. 4.6). Thus

$$Q_G \approx -Q_B - Q_o \quad (7.4.34)$$

To find Q_I accurately, we can use the observation in Sec. 4.6 that Q'_I varies as a straight line with position between its value at the source and its value at the drain, as shown in Fig. 4.24. Thus

$$Q'_I(x) = Q'_{I0} + \frac{x}{L} (Q'_{IL} - Q'_{I0}) \quad (7.4.35)$$

where expressions for Q'_{I0} and Q'_{IL} have been given in Sec. 4.6. Since we have Q'_I as an explicit function of x , we can use it in (7.2.3a) directly to find Q_I . The result is (Prob. 7.7)

$$Q_I = WL \frac{Q'_{I0} + Q'_{IL}}{2} \quad (7.4.36)$$

Similarly, from (7.3.9) and (7.4.35), we obtain (Prob. 7.7)

$$Q_D = WL \left(\frac{Q'_{I0}}{6} + \frac{Q'_{IL}}{3} \right) \quad (7.4.37)$$

$$Q_S = WL \left(\frac{Q'_{I0}}{3} + \frac{Q'_{IL}}{6} \right) \quad (7.4.38)$$

In practice, the above three charges are often neglected completely in the computation of transients for the following reason. The source and drain junction depletion regions (in the *extrinsic* part of the device, outside the channel area) contain charges which can be evaluated as in Sec. 1.5. Of those charges the ones in the n^+ material must be changed (when V_{SB} and V_{DB} are varied) by electrons supplied through the source and drain terminals. These charges and the corresponding charging currents are much larger than the ones associated with the inversion layer for typical channel lengths. Thus, in weak inversion, the source and drain charging currents are dominated by the extrinsic part of the device, and the following simplification is often used:

$$Q_I \approx Q_D \approx Q_S \approx 0 \quad (7.4.39)$$

This should not be interpreted to mean that the weak-inversion layer is irrelevant in dynamic operation. In fact, it turns out that in weak inversion speed is limited by non-quasi-static effects in the inversion layer, rather than by extrinsic parasitics.

7.4.5 General Charge Sheet Model

We have seen in Secs. 7.4.2 and 7.4.4 that, if strong inversion or weak inversion is considered separately, simple expressions become possible for the charges in each of the two regions. However, it is also possible to develop a general expression for each charge which will be valid in *all* regions of inversion, just as was done for the drain current in Sec. 4.3. These expressions will even be valid in moderate inversion, a region for which no simple expression has been derived. As usual, extra complexity must be accepted to make such generality possible.

The principle on which the derivation of general charge expressions is based is simple. The general expression for the inversion layer charge per unit area, as used in Sec. 4.3.1, is

$$Q'_I = -C'_{ox} \left(V_{GB} - V_{FB} - \psi_s + \frac{Q'_B}{C'_{ox}} \right) \quad (7.4.40)$$

where the corresponding general expression for the depletion region charge per unit area is

$$Q'_B = -\gamma C'_{ox} \sqrt{\psi_s} \quad (7.4.41)$$

Equation (4.3.5) gives the drain-to-source current, assuming both drift and diffusion are present. From this equation we have

$$dx = -\frac{\mu W}{I_{DS}} Q'_I d\psi_s + \frac{\mu W}{I_{DS}} \phi_t dQ'_I \quad (7.4.42)$$

Appropriate use of the above three equations in (7.2.3) and (7.3.9) results in general expressions for the total charges Q_I , Q_B , Q_G , Q_D , and Q_S .† As an example, by using (7.4.42) in (7.2.3a), we obtain

$$Q_I = -\frac{W^2 \mu}{I_{DS}} \int_{\psi_{s0}}^{\psi_{sL}} Q_I'^2 d\psi_s + \frac{W^2 \mu}{I_{DS}} \phi_t \int_{Q'_{I0}}^{Q'_{IL}} Q'_I dQ'_I \quad (7.4.43a)$$

or

$$Q_I = -\frac{W^2 \mu}{I_{DS}} \int_{\psi_{s0}}^{\psi_{sL}} Q_I'^2 d\psi_s + \frac{W^2 \mu}{I_{DS}} \phi_t \frac{1}{2} (Q_{IL}'^2 - Q_{I0}'^2) \quad (7.4.43b)$$

where ψ_{s0} and ψ_{sL} are the surface potentials at the source and the drain ends of the channel, respectively, and I_{DS} is found as in Sec. 4.3. The first integral in (7.4.43b) can be evaluated after expressing Q'_I in terms of ψ_s , using (7.4.40) and (7.4.41). The result will be in terms of ψ_{s0} and ψ_{sL} , which can be determined from (4.3.18). The quantity in parentheses in (7.4.43b) can similarly be found by expressing Q'_{IL} in terms of ψ_{sL} and Q'_{I0} in terms of ψ_{s0} . Note that ψ_{sL} and ψ_{s0} must be *very* accurately known, as discussed in Sec. 4.3.1.

As a check of (7.4.43b) consider the special case of strong inversion in which only drift current is assumed. Then, only the first term will be present in the right-hand side of (7.4.42) (Sec. 4.3). Thus, only the term containing the integral will be present in (7.4.43b). By using in it (4.5.1) and (4.5.4), this term is seen to reduce to (7.4.2c), which was developed specifically for strong inversion. If now weak-inversion operation is considered instead, the current is practically all due to diffusion and only the second term will be present in (7.4.42). Thus, only the second term will appear in (7.4.43b). By using in it I_{DS} from (4.6.3), this term is seen to reduce to (7.4.36), which was developed specifically for weak inversion. In moderate inversion, both terms in (7.4.43b) will be significant.

The rest of the charges can be similarly found, and can be expressed in terms of ψ_{s0} and ψ_{sL} . The procedure is rather long but mathematically straightforward and involves changes of variables and integration (Prob. 7.8).‡

†A different, more complicated procedure involves the quasi-Fermi potential in the channel.²⁸

‡We note that, for implementation in a computer program, these expressions may have to be modified since they are not numerically robust. For example, in (7.4.43b), as ψ_{sL} approaches ψ_{s0} , both the numerator and the denominator approach 0.

Alternatively, the expression in (7.4.43b) can be evaluated in a manner consistent with the simplified charge sheet models of Sec. 4.3.2 (Prob. 7.9). Consider, for example, the model in (4.3.41b), repeated here:^{57,61}

$$I_{DS} = \frac{W}{L} \mu \left[\frac{-1}{2nC'_{ox}} (Q'_{IL}{}^2 - Q'_{I0}{}^2) + \phi_t (Q'_{IL} - Q'_{I0}) \right] \quad (7.4.44)$$

The development of this model in Sec. 4.3.2 can be traced to the observation⁴⁷ that, for a given V_{GB} , Q_I varies almost linearly with ψ_s . This resulted in (4.3.27) which, from (4.3.38), can be written as follows:

$$\frac{dQ'_I}{d\psi_s} = nC'_{ox} \quad (7.4.45)$$

This relation can be used to change the variable of integration in (7.4.43b) from ψ_s to Q'_I . This results in the following expressions for the total charge Q_I :

$$Q_I = \frac{W^2 \mu}{I_{DS}} \left[-\frac{Q'_{IL}{}^3 - Q'_{I0}{}^3}{3nC'_{ox}} + \frac{\phi_t}{2} (Q'_{IL}{}^2 - Q'_{I0}{}^2) \right] \quad (7.4.46)$$

Using (7.4.44) in this expression and removing the common factor $(Q'_{IL} - Q'_{I0})$ from both numerator and denominator, we obtain:^{57,61}

$$Q_I = WL \frac{\frac{2}{3} (Q'_{IL}{}^2 + Q'_{IL}Q'_{I0} + Q'_{I0}{}^2) - n\phi_t C'_{ox} (Q'_{IL} + Q'_{I0})}{Q'_{IL} + Q'_{I0} - 2n\phi_t C'_{ox}} \quad (7.4.47)$$

which can be evaluated for known bias voltages, by relating Q'_{IL} to ψ_{sL} and Q'_{I0} to ψ_{s0} , as before. This expression gives Q_I in all regions of inversion. As a check, consider strong inversion and operation in saturation. Then the last terms in the numerator and denominator of (7.4.47) (the development of which terms can be traced to the assumption that diffusion currents are present) will be negligible, and Q'_{IL} will be 0; thus (7.4.47) gives $Q_I = \frac{2}{3}(W/L)Q'_{I0}$. With the help of (4.5.10), this is easily seen to reduce to (7.4.27).

Expressions for the other charges can similarly be developed for the simplified charge sheet model; the interested reader is referred to the literature.^{57,61} Continuous expressions have also been proposed using interpolation semiempirical models.⁵⁸

7.4.6 Depletion

In digital circuits, transistors are switched between conduction and cutoff. Therefore, the charges in the latter state are of importance in calculating the transient response of such circuits. The cutoff region consists of two regions—depletion and accumulation (Fig. 3.2). We consider these regions here and in the following subsection. In the depletion region, the inversion layer charge is totally negligible (Chap. 2):

$$Q_I = 0 \quad (7.4.48)$$

Because of this, the development of expressions for Q_B and Q_G is identical to that used for weak inversion in Sec. 7.4.4 and results in the same equations, repeated here for convenience:

$$Q_B = -WLC'_{ox}\gamma\left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}}\right) \quad (7.4.49)$$

$$Q_G = -Q_B - Q_o \quad (7.4.50)$$

7.4.7 Accumulation

As explained in Chap. 2, when V_{GB} is sufficiently less than V_{FB} , holes in the p substrate (where they are in abundance) accumulate immediately below the oxide and form a very thin, highly conductive sheet there. The surface potential needed to support these holes is negative but of very small magnitude because the sheet of holes is so thin (this can be verified using the general analysis in Appendix I). Hence, ψ_s can be neglected in the potential balance equation (2.3.1), which results in an oxide potential $\psi_{ox} = V_{GB} - \phi_{MS}$. The gate charge per unit area $Q'_G = C'_{ox}\psi_{ox}$ is thus known, and multiplied by the gate area gives

$$Q_G = WLC'_{ox}(V_{GB} - \phi_{MS}) \quad (7.4.51)$$

The charge of the holes in the bulk, denoted by Q_C , can now be found from the charge balance equation $Q_G + Q_C + Q_o = 0$ [see (2.3.3)]. Thus

$$Q_C = -Q_G - Q_o \quad (7.4.52)$$

In the literature, the presence of Q_o in (7.4.52) is sometimes overlooked. The corresponding equation for Q_G in such cases uses V_{FB} rather than the correct ϕ_{MS} as above.

The accuracy of (7.4.51) decreases somewhat if V_{GB} is very close to V_{FB} , since then the accumulation is not heavy and the arguments used above do not really hold. If high accuracy is desired, one can use the material of Appendix I.

7.4.8 Plots of Charges versus V_{GS}

In Fig. 7.7, we have seen the charges plotted as a function of V_{DS} , for a fixed V_{GS} in strong inversion. To show the behavior of the charges in all regions of operation, the opposite is often done. For a fixed V_{DS} , charges are plotted vs. V_{GS} . This corresponds to moving up a vertical line in the I_{DS} - V_{DS} characteristics, as shown in Fig. 7.8. The regions encountered as V_{GS} is increased are accumulation, depletion, weak inversion, moderate inversion, saturation, and nonsaturation, in that order. Recall that, for the approximate strong-inversion model, nonsaturation is defined by

$$V_{DS} \leq \frac{V_{GS} - V_T}{\alpha} \quad (7.4.53)$$

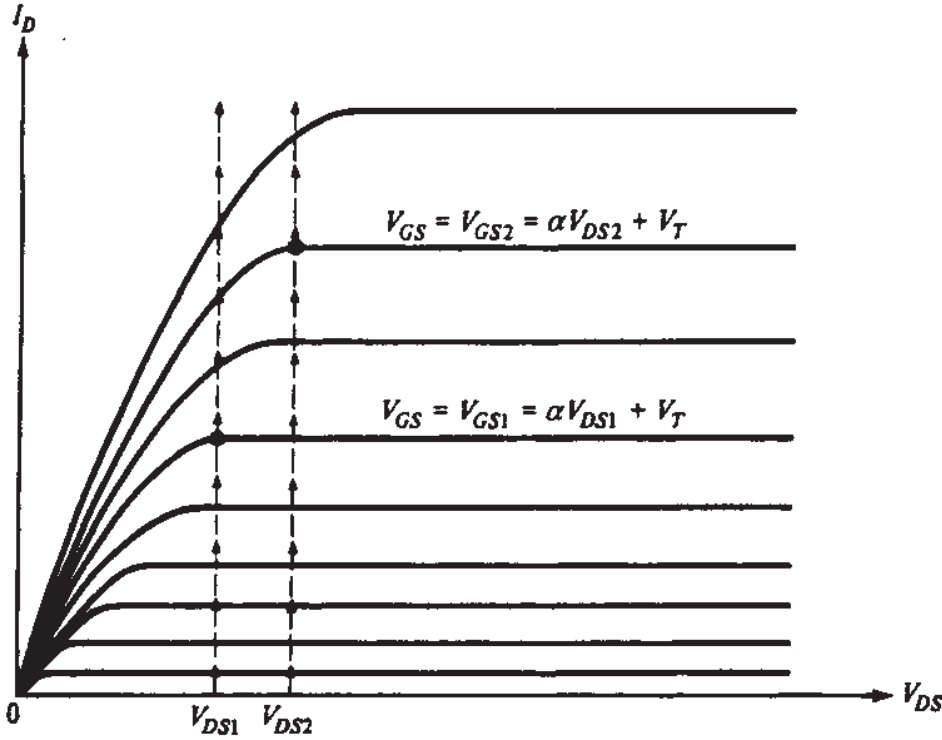


FIGURE 7.8

Drain current versus drain-source voltage, with gate-source voltage as a parameter. V_{DS1} and V_{DS2} are two drain-source voltage values for which the plots of Fig. 7.9 are obtained.

which, solved for V_{GS} , gives

$$V_{GS} \geq \alpha V_{DS} + V_T \quad (7.4.54)$$

The critical value in the right-hand side of this equation is marked in Fig. 7.8. The charges vs. V_{GS} for the fixed value of $V_{DS} = V_{DS1}$ shown in Fig. 7.8 are plotted with solid lines in Fig. 7.9. (The middle plot gives the depletion region charge Q_B in inversion and depletion, and the charge of holes Q_C in accumulation.) If V_{DS} were chosen at a value $V_{DS2} > V_{DS1}$ (Fig. 7.8), we would obtain the broken lines instead. Note that V_{DS} makes a difference only in strong inversion nonsaturation in these plots. This is because, as seen in the charge expressions developed in this section, in accumulation and in depletion V_{DS} does not appear at all in them, and in weak inversion, moderate inversion, and strong-inversion saturation we are in the flat part of the characteristics in Fig. 7.8 where the drain cannot control the channel.

7.4.9 Use of Charges in Evaluating the Terminal Currents

The expressions we have developed above for the charges in individual regions of operation can be written as functions of V_D , V_G , V_B , and V_S by substituting in them $V_{GS} = V_G - V_S$, $V_{DS} = V_D - V_S$, and $V_{SB} = V_S - V_B$. As follows from the comments in

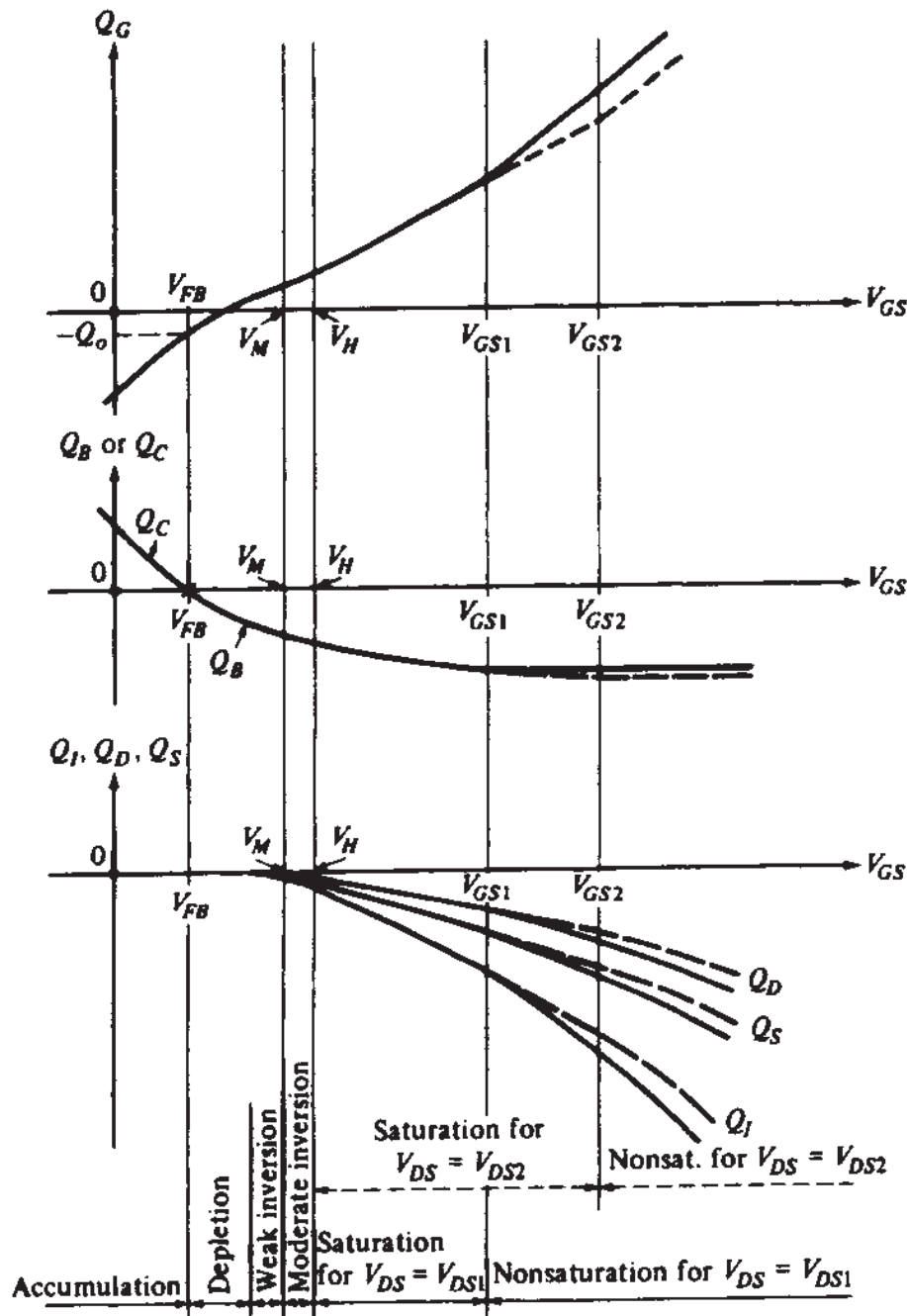


FIGURE 7.9

Total charges versus gate-source voltage for two values of V_{DS} . Solid line: $V_{DS} = V_{DS1}$; broken line: $V_{DS} = V_{DS2}$, $V_{DS2} > V_{DS1}$ (see Fig. 7.8).

Sec. 7.4.1, if the terminal voltages vary slowly enough so that quasi-static operation is maintained, the same expressions can be used to determine the time-varying charges. Thus the partial derivatives in (7.3.16) can be determined. In other words, we have

$$\frac{\partial q_K}{\partial v_L} = \left. \frac{\partial Q_K}{\partial V_L} \right|_{V_L = v_L}, \quad \text{quasi-static operation} \quad (7.4.55)$$

where each of K and L can stand for D , G , B , or S . Thus, if the variation of terminal voltages with time is known, the terminal currents can be found from (7.3.16), (7.4.55), (7.3.4), and (7.3.12).

A similar approach can be taken if the general charge sheet model is used (Sec. 7.4.5), since the surface potentials ψ_{s0} and ψ_{sL} can be related to V_{SB} and V_{DB} , respectively, through (4.3.18). The algebra, however, is considerably more complicated.

7.5 TRANSIT TIME UNDER DC CONDITIONS

The transit time (Sec. 1.3.1) associated with dc operation in a transistor is the average time it takes for an electron to travel the length of the channel:

$$\tau = \frac{|Q_I|}{I_{DS}} \quad (7.5.1)$$

Having evaluated the inversion layer charge in the previous section, this is a convenient time to calculate τ for four cases of interest. We will make use of τ in the following section.

1. *Strong-inversion nonsaturation with very small V_{DS} .* From (7.4.22) we have $|Q_I| \approx C'_{ox} WL(V_{GS} - V_T)$, and from (4.5.37a), with very small V_{DS} , $I_{DS} \approx \mu C'_{ox}(W/L)(V_{GS} - V_T)V_{DS}$. Hence, from (7.5.1),

$$\tau \approx \frac{L^2}{\mu V_{DS}} \quad (7.5.2)$$

Note that, since with negligible V_{DS} the channel is approximately uniform and the drift velocity approximately fixed, we could have used (1.3.10) to obtain the above result.

2. *Strong-inversion saturation.* Assume that no velocity saturation takes place. From (7.4.27), $|Q_I| = \frac{2}{3}WLC'_{ox}(V_{GS} - V_T)$, and from (4.5.37b), $I_{DS} = \frac{1}{2}\mu C'_{ox}(W/L)(V_{GS} - V_T)^2/\alpha$. We thus obtain, from (7.5.1),

$$\tau = \frac{4}{3}\tau_o \quad (7.5.3)$$

where

$$\tau_o = \frac{\alpha L^2}{\mu(V_{GS} - V_T)} \quad (7.5.4)$$

For a device with $\mu = 600 \text{ cm}^2/(\text{V}\cdot\text{s})$ and $L = 1 \text{ }\mu\text{m}$, operated with $(V_{GS} - V_T) = 2 \text{ V}$ and with $\alpha = 1.2$, the value of τ is 13 ps.

3. *Weak inversion with $V_{DS} > 5\phi_t$.* Here, from Sec. 4.6 we easily see that $Q'_{IL} \approx 0$, and from (7.4.36), $|Q_I| = \frac{1}{2} |Q'_{I0}| WL$. From (4.6.12), $I_{DS} = \mu(W/L)\phi_t |Q'_{I0}|$. Hence

$$\tau = \frac{L^2}{\mu(2\phi_t)} \quad (7.5.5)$$

With $\mu = 600 \text{ cm}^2/(\text{V}\cdot\text{s})$ and $L = 1 \text{ }\mu\text{m}$, τ is about 320 ps. Note that in all three cases, the transit time is proportional to the *square* of L . This, of course, is so because Q_I is proportional to L , and I_{DS} is inversely proportional to L . To see this effect another way, consider the case of nonsaturation with very small V_{DS} . The channel is nearly uniform and the field is everywhere approximately equal to V_{DS}/L . The drift velocity of the electrons is proportional to this field. Increasing L , say, m times decreases the drift velocity m times and these electrons now have to travel m times the distance. So the time it takes them to travel the length of the channel will increase m^2 times. Applying the simple model of Sec. 1.3.3 to weak inversion, a similar argument holds, only there the “driving force” is not the field but rather the gradient of the charge in the channel, $|Q'_{I0}|/L$.

4. *Velocity saturation.* If velocity saturation is present over part of the channel, the above arguments break down.⁴⁰ Although the value of the transit time in this case can be calculated by using material from Sec. 6.5, we will limit our discussion to a simple estimate. We note that τ will be larger than what one would have if the electrons were moving at maximum speed over *all* of the channel. Thus

$$\tau > \frac{L}{|v_d|_{\max}} \quad (7.5.6)$$

Figure 7.10 shows the transit time of a device operating in the “flat” part of the I_{DS} - V_{DS} characteristics as a function of V_{GS} (solid line). As V_{GS} is increased V'_{DS} increases (Sec. 4.5.3) and V_{DS} must be raised if the device is to be kept in the saturation region. The increase in V'_{DS} is accompanied by an increased electric field and, thus, if L is small, velocity saturation sets in over much of the channel. Thus, τ cannot be decreased at will by increasing V_{GS} , as one could have concluded by carelessly applying (7.5.3) and (7.5.4).

7.6 LIMITATIONS OF THE QUASI-STATIC MODEL

From our discussion so far, we expect that the quasi-static model will be valid if the terminal voltages vary sufficiently slowly so that the charge distribution in the channel can follow with negligible inertia. A quantitative definition of the term “suffi-

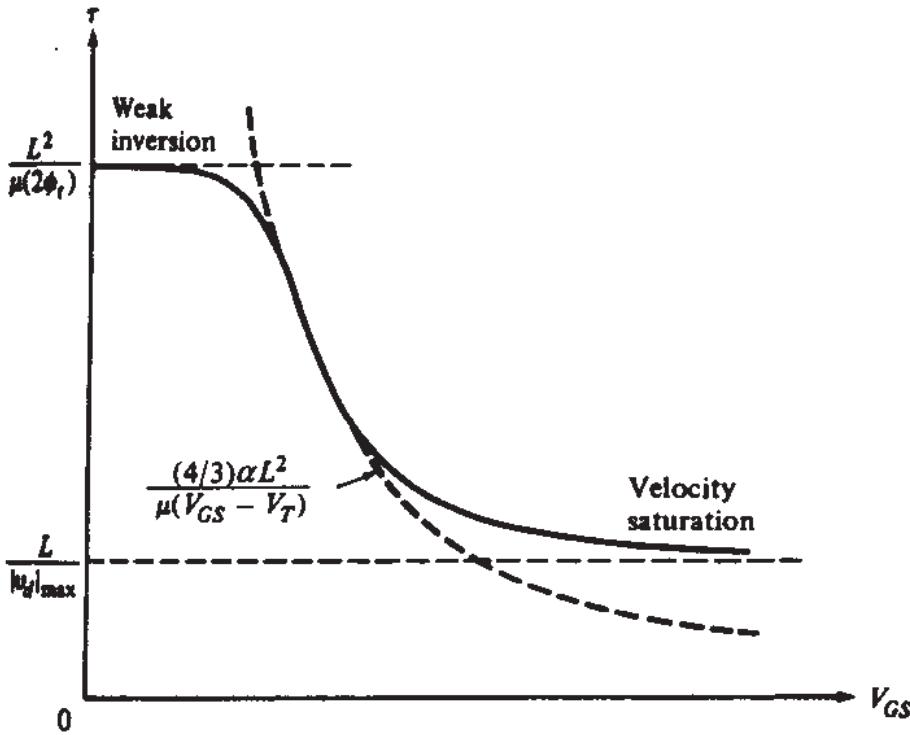


FIGURE 7.10

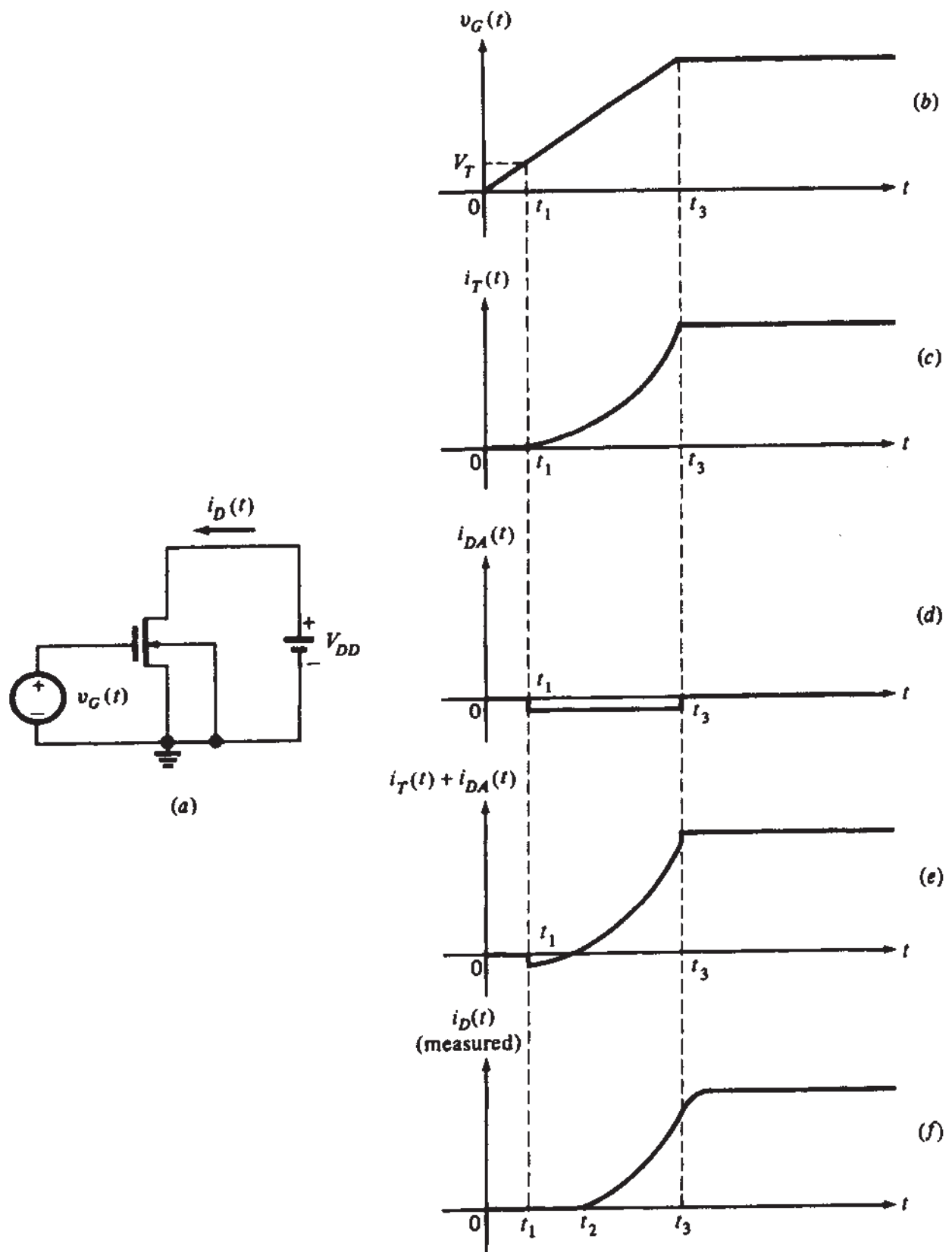
Transit time versus gate-source voltage for operation in the “flat” part of the I_{DS} - V_{DS} characteristics.

ciently slowly” is difficult to come by. Whether results obtained from using the quasi-static model are trustworthy or not depends on the type of voltage waveforms applied to the terminals, on the regions of operation involved, on the type of result desired (current waveform shape, delay, risetime, etc.), on the accuracy sought, etc. In practice, some rules of thumb have been developed semiempirically, by using a simple basic case shown in Fig. 7.11a. Here only v_G is varying, as shown in Fig. 7.11b. We will consider only the intrinsic device effects. The drain current contains a transport part $i_T(t)$ and a charging part $i_{DA}(t)$, as in (7.3.4a):

$$i_D(t) = i_T(t) + i_{DA}(t) \quad (7.6.1)$$

For approximate calculations in digital circuit applications, it is often assumed that a transistor is in the off state if $v_{GS} < V_T$ and in strong inversion if $v_{GS} > V_T$. Using this simplification here implies that the device goes abruptly from off to saturation operation at $t = t_1$. We will also assume that V_{DD} is large enough so that the device never goes into nonsaturation. The conductive part of the current can be found from any dc model, such as (4.5.37b), and is shown in Fig. 7.11c. The charging current $i_{DA}(t)$ can be found from (7.3.16a). Since v_D , v_S , and v_B are constant, this equation gives

$$i_{DA}(t) = \frac{\partial q_D}{\partial v_G} \frac{dv_G}{dt} \quad (7.6.2)$$

**FIGURE 7.11**

(a) A transistor with varying excitation; (b) gate-source voltage; (c) transport current calculated from a dc model; (d) drain charging current calculated by assuming quasi-static operation; (e) sum of (c) and (d); (f) form of actual $i_D(t)$ observed in practice. A sufficiently large V_{DD} is assumed, so that even the maximum value of v_G corresponds to operation in the saturation region. Extrinsic effects are assumed negligible.

In saturation, the value of q_D in quasi-static operation $Q_{D,\text{sat}}$ is given by (7.4.28). Thus, $\partial q_D / \partial v_G$ is $-\frac{4}{15} WLC'_{\text{ox}}$, a negative constant independent of v_G . (In a more general case, though, one can expect the partial derivatives $\partial q_k / \partial v_l$ to be functions of the terminal voltages.) Since dv_G/dt is constant for the rising part of v_G , $i_{DA}(t)$ is of the form shown in Fig. 7.11d. Adding the two drain current components then produces $i_D(t)$, as shown in Fig. 7.11e.

Measurements of $i_D(t)$ reveal a waveform of the type shown in Fig. 7.11f assuming negligible extrinsic effects.¹⁹⁻²¹ This is seen to differ from our “quasi-static” result of Fig. 7.11e, notably in two respects:

1. For some time after t_1 , the quasi-static model predicts a negative drain current, which is not observed in practice *once extrinsic device effects are subtracted out* (Prob. 7.11). Instead, the drain current remains at zero until some time t_2 , as shown in Fig. 7.11f. This can be explained as follows. For $t < t_1$, the channel is empty. At $t = t_1$, the conditions in the channel become favorable for electrons; the latter enter the channel through the source and move toward the drain. However, no drain current will be observed until the electrons *reach* the drain. This happens at time t_2 , as shown in Fig. 7.11f. In contrast to this picture, the quasi-static model assumes that at *any* time t' after t_1 the channel contains electrons *throughout* its length, as would be the case if v_G had been frozen for a long time at the value $v_G(t')$, i.e., the nonzero time it takes for the electrons to reach the drain is not considered.
2. At $t = t_3$, $i_D(t)$ in Fig. 7.11e is predicted to jump to its dc steady-state value instantaneously. This is a consequence of the assumption that the charge adjusts itself with no inertia. In reality this is not the case, as illustrated in Fig. 7.11f.

The quasi-static model is thus seen to fail when it comes to predicting the fine details of the drain current waveform. However, for many applications this is of no great consequence, since the fine waveform details are often of no great interest. This is, for example, the case in much of digital circuit design, where quantities of interest are initial values, final values, output rise or fall times, delays between output and input, etc. Comparisons to measurements and to non-quasi-static numerical results have been used to suggest that, for much of digital circuit work, the quasi-static model can be used with acceptable results if the rise time t_R of the waveforms† involved satisfies the condition^{20,21}

$$t_R > 20\tau_o \quad (7.6.3)$$

where τ_o is given by (7.5.4) by using the maximum value of V_{GS} from Fig. 7.11b, assuming no velocity saturation is present. It is emphasized that the above is only a

†Defined for this purpose as $t_3 - t_1$ in Fig. 7.11b.

rough rule of thumb.† For example, depending on the application, the factor of 20 could instead be 15 or 25. As a numerical example of the order of magnitude involved here, consider a device with $\mu = 600 \text{ cm}^2/(\text{V}\cdot\text{s})$, $V_T = 0.5 \text{ V}$, $L = 1 \text{ }\mu\text{m}$, $\alpha = 1.2$, and $V_{GS,\text{max}} = 3 \text{ V}$. Then (7.6.3) requires $t_R > 160 \text{ ps}$.

In common fabrication technologies, in which the transistor exhibits significant extrinsic parasitic capacitances, the above limit is not restrictive. The speed of operation is slowed down in these technologies because of such “stray” capacitances, and internal waveforms on a chip will often have rise times larger than this limit. On top of this, even if occasionally a gate voltage waveform has a rise time somewhat below this limit, the total transient currents and delays due to the stray capacitance of the device being driven can be significant and can mask the errors due to intrinsic effects predicted by quasi-static models. The above is not true with technologies which achieve very low stray capacitance values, such as SOI technologies (Sec. 1.6). In such cases, as well as in cases where pushing the speed limit is attempted by using bulk technologies, results obtained using quasi-static models should be looked at with suspicion.

THE ISSUE OF DRAIN/SOURCE CHARGE PARTITION. We have seen that q_D and q_S can be evaluated from (7.3.9a) and (7.3.9b). These relations satisfy $q_D + q_S = q_I$, but the ratio q_D/q_S depends on the bias voltages. Consider, for example, operation in strong-inversion saturation. Then the above relations lead to (7.4.28) and (7.4.29). Comparing these to (7.4.27), we see that q_D is 40 percent, and q_S 60 percent, of the total charge q_I , which is often referred to as a “40/60 partition.” Other partitions have also been proposed. One of these is 50/50, which is clearly not appropriate for the nonsymmetric situation that exists in the saturation region. Another partition used is 0/100, which means $q_S = q_I$ and $q_D = 0$. However, the claim that $q_D = 0$ has been disputed in the two paragraphs following (7.4.30). This claim is sometimes defended by pointing to the fact that, if q_D is taken identically zero, $i_{DA}(t)$ from (7.6.2) will be zero, too; thus no negative current region will occur in $i_D(t)$ plots, which agrees with measurements, unlike the case in Fig. 7.11e, which was obtained using a 40/60 partition. Note, however, four things:

1. It is the negative i_{DA} that makes the positive part of the $i_D(t)$ plot in Fig. 7.11e resemble what is measured in practice (compare to Fig. 7.11f). If q_D , and thus i_{DA} , were zero, i_D would be identical to i_T in Fig. 7.11c, which is not necessarily an improvement. Thus, in Fig. 7.11, which of the two waveforms in c and e is a better approximation to f will depend on the application.
2. Although it is true that in the case of Fig. 7.11a no negative i_D would normally be observed, this would not necessarily be the case if V_{DD} were lower. Indeed, in non-saturation negative drain current transients are possible, and this has been verified

†The origin of this result will be understood after non-quasi-static analysis is described in Sec. 7.7.

by measurements.[†] Such currents would not be possible to predict if q_D were taken to be identically zero.

3. A 0/100 choice may ensure that i_D flows in only one direction, but it does not ensure this for i_S .
4. Notice that i_{DA} depends on the rate of change of v_G , as seen in (7.6.2). It can be shown that if (7.6.3) is satisfied, the negative excursion in the i_D plot is limited to only a couple of percent of the eventual, maximum value of i_D . Equation (7.6.2) will only predict a negative current with a large magnitude if dv_G/dt is large, in which (7.6.3) is violated. However, in such cases the quasi-static model should not be used in the first place!

Thus, to conclude, we can state the following: *for rates of change at which the assumption of quasi-static operation is valid*, (7.3.9) is the correct choice for q_D and q_S . Several facts in support of this claim have been listed following (7.3.12).

MULTISEGMENT MODELS. One way to model a transistor at speeds where its quasi-static model breaks down is to view it as consisting of several sections, each section being short enough to be modeled quasi-statically. This idea is shown in Fig. 7.12. In *b*, each box is assumed to be a “subtransistor,” with its own imaginary source and drain *points*. Of course, with the exception of the left and right extremes, these subtransistors are assumed to consist only of intrinsic parts. In other words, no extrinsic source and drain *regions* at intermediate points are assumed to exist.

OTHER EFFECTS. Note that in our discussion above we have not considered short-channel effects. The latter can greatly complicate the picture.^{19–21,38,45–53,55,56} Among the phenomena observed in them, but not in truly long-channel devices, are a *transient transport* current²¹ (in addition to the transient charging current) and, of course, velocity saturation which renders (7.4.1) and the relations based on it invalid. Two-dimensional numerical simulation is a valuable tool in this case.^{19,20} Finally, there is another phenomenon, not related to short-channel effects, that we have not considered in our modeling. This phenomenon is observed during the falling part of a gate voltage waveform. As $v_G(t)$ is decreased, the magnitude of the inversion layer charge must be reduced. To this end, electrons exit through the drain and source terminals. The dynamics of this removal process show that there is a finite removal capability associated with it.⁷ If v_G is decreased too fast, the ensuing “bottlenecks” at either end

[†]If V_{DD} were smaller, so that the maximum value of v_G corresponded to the *nonsaturation* region, we would have a situation analogous that in Fig. 1.25c. Here, as the piston moved downward fast and stopped at the position shown, fluid would initially flow into the “channel” not only from the source but also from the drain. This would correspond to a *negative* drain current. After things settled, of course, a regular flow toward the right would have to be established, so the drain current would go through zero and then would become positive. These predictions are, indeed, verified by actual measurements on transistors.

of the channel result in some electrons being temporarily “trapped” in the channel. The field-induced junction consisting of the inversion layer and the substrate can then become momentarily forward-biased and the electrons can cross into the bulk. There they recombine with holes and cause a substrate current to flow. This phenomenon is called *charge pumping*. It is more pronounced for shorter falling times of the gate voltage waveform, as might be expected intuitively.^{20,59} It is estimated that 1 percent of the total inversion layer charge will exit through the substrate if the fall time is 0.04 ns for a 3- μm -long device, or about 1 ns for a 10- μm -long device.²⁰

We conclude this section by reminding the reader that parasitic elements associated with the *extrinsic* part of the device can significantly alter the behavior of the transistor, compared to that predicted here. Such elements include the gate-source and gate-drain overlap capacitances, and the substrate-source and substrate-drain junction capacitances. In addition, the fact that the substrate is not a perfect conductor can become important. The significant resistance associated with the latter, in conjunction with the intrinsic and extrinsic substrate capacitances, can affect the dynamic performance of the device. Extrinsic elements will be considered in Sec. 8.4.

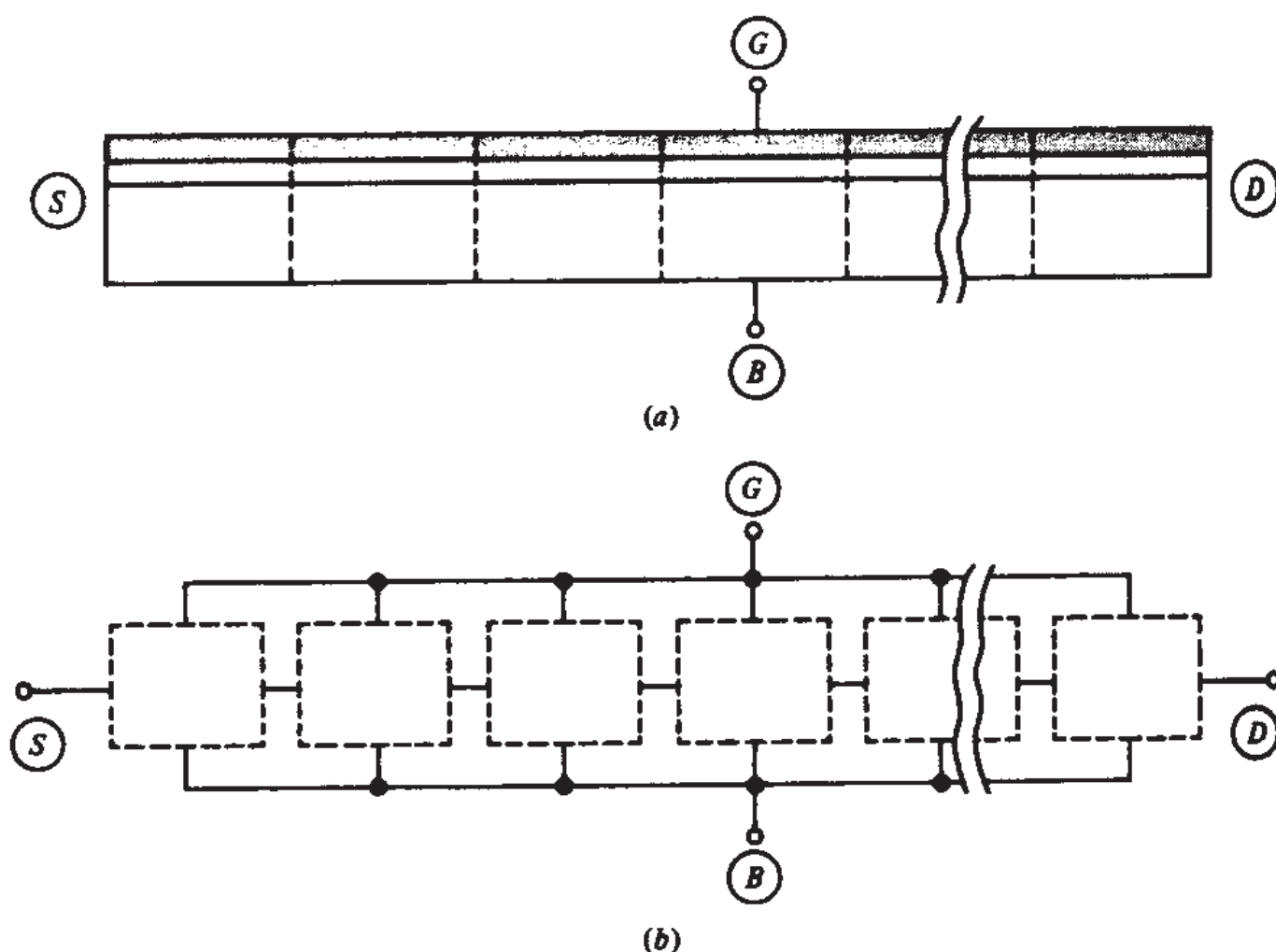


FIGURE 7.12

(a) A long-channel transistor separated into several sections; (b) model for (a); each box represents a quasi-static model corresponding to one section in (a).

7.7 NON-QUASI-STATIC MODELING

7.7.1 Introduction

In the previous section it was seen that, for a given channel length, the quasi-static model breaks down if the input changes too fast. It was suggested that one way to extend the validity of quasi-static models in that case is to consider the device as a connection of several shorter devices (Fig. 7.12) and to model each section quasi-statically. Note that, for each section, the current entering one end is, in general, different from the current leaving the other end during transients. This accounts for possible inversion layer charge buildup within each section, just as different magnitudes of drain and source current are encountered because of charge buildup in the single-section quasi-static model. The faster the change of the input, the shorter each section must be, and the larger the number of sections. In the limit, one can let the section length approach zero, thus making the total number of sections approach infinity. The resulting model would then not be subjected to the speed limitations of quasi-static models. We will develop this idea formally in this section.^{1-4,6,9,12,13,17,19,20,22,32,54,62-69,71} Note that, following the argument above concerning currents, we should allow for the current to be a function of position x along the channel as well as, of course, time:

$$i = i(x, t) \quad (7.7.1)$$

Similarly, the inversion layer charge per unit area will be a function of position and time:

$$q'_I = q'_I(x, t) \quad (7.7.2)$$

7.7.2 The Continuity Equation

Consider a section of the inversion layer of very small length Δx , as shown in Fig. 7.13. Let the current entering on the right be in general different from that leaving from the left by a small amount Δi , as shown. The total charge entering on the right

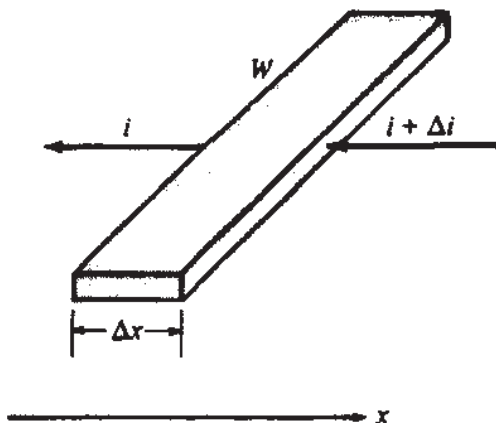


FIGURE 7.13

A chunk of the inversion layer of small length Δx .

in a small time interval Δt is $(i + \Delta i)\Delta t$; the total charge leaving from the left in the same amount of time is $i\Delta t$. Thus, in the interval Δt , the charge inside the chunk must be increasing by $(i + \Delta i)\Delta t - i\Delta t = \Delta i \Delta t$. The corresponding increase $\Delta q'_I$ in the inversion charge *per unit area* will simply be the total charge increase divided by the chunk's area, as seen from above. Thus,

$$\Delta q'_I = \frac{\Delta i \Delta t}{W \Delta x} \quad (7.7.3)$$

which can also be written as

$$\frac{\Delta i}{\Delta x} = W \frac{\Delta q'_I}{\Delta t} \quad (7.7.4)$$

We now let the finite differences approach 0. The left-hand side then becomes the partial derivative of $i(x, t)$ with respect to x . Similarly, the fraction on the right becomes the partial derivative of $q'_I(x, t)$ with respect to t . Thus,

$$\boxed{\frac{\partial i(x, t)}{\partial x} = W \frac{\partial q'_I(x, t)}{\partial t}} \quad (7.7.5)$$

This equation is referred to as the *continuity equation*.† It is simply a way to express charge conservation for a chunk of infinitesimal length. Note that if q'_I does not change with time ($\partial q'_I / \partial t = 0$), the above equation gives $\partial i / \partial x = 0$, that is, i then has a constant value independent of position x . This is because there is no charge buildup, and thus the current exiting must be equal to the current entering. That is the case under dc conditions. In fact, the constancy of i was instrumental in developing the dc I_D equations in Chap. 4.

7.7.3 Non-Quasi-Static Analysis

Non-quasi-static analysis of the MOS transistor is a difficult mathematical exercise.^{2,6,12,13,17,22,32,54,62-69,71} We will illustrate it for the special case in which all points in the channel are in *strong inversion*, which simplifies matters considerably. We can relate $q'_I(x, t)$ to the external terminal voltages and the internal effective reverse bias $V_{CB}(x, t)$ by the time-varying version of (4.5.10a):

$$q'_I(x, t) = -C'_{ox} \left[v_{GB}(t) - V_{FB} - \phi_0 - v_{CB}(x, t) - \gamma \sqrt{\phi_0 + v_{CB}(x, t)} \right] \quad (7.7.6a)$$

†The reader may have encountered this equation with i defined in the opposite direction from that in Fig. 7.13, in which case a minus sign would appear in one side of the equation. The above form of the continuity equation is appropriate for our purposes. Note also that, while in the case we are considering only electrons are assumed to be present as usual, in more general cases the continuity equation remains valid if the current and charge in it are taken to include the effect of both electrons and holes. If separate equations are written for each carrier, additional terms must be included to account for carrier generation and recombination.³³

The variation of $v_{CB}(x, t)$ with x is the “driving force” for the current flow. This is expressed from (4.5.6) with, of course, I_{DS} replaced by $i(x, t)$:

$$i(x, t) = -\mu W q'_l(x, t) \frac{\partial v_{CB}(x, t)}{\partial x} \quad (7.7.6b)$$

Finally, the continuity equation developed in the previous subsection is

$$\frac{\partial i(x, t)}{\partial x} = W \frac{\partial q'_l(x, t)}{\partial t} \quad (7.7.6c)$$

Equations (7.7.6) constitute a system of three equations in three unknowns: $q'_l(x, t)$, $i(x, t)$, and $v_{CB}(x, t)$.† The last two equations express basic facts about current flow. They have both been developed by considering a chunk of material with mobile electrons in it. Apart from the convenient symbols used and the assumption of only drift current in (7.7.6b), the development was independent of whether the material is part of an MOS structure. The MOS transistor physics only enters in the first equation.

The solution of (7.7.6) requires a set of *initial and boundary conditions*. These will depend on the terminal voltages. As an example, consider the circuit of Fig. 7.14a with the step input shown in Fig. 7.14b.⁶ The device is assumed to have settled in the off condition before the positive step is applied. The values of the step V and of V_{DD} are such that, after a transient period, the device settles in the saturation region. Since the device is initially off, q'_l will be zero at $t = 0$ everywhere in the channel:

$$q'_l(x, 0) = 0 \quad (7.7.7a)$$

With the large value $v_{GB}(t) = V$ applied at $t = 0$, the source end of the channel ($x = 0$) is assumed to reach strong inversion immediately after $t = 0$. Since at that end $v_{CB} = v_{SB} = 0$, (7.7.6a) gives, for all positive time

$$q'_l(0, t) = -C'_{ox} (V - V_{FB} - \phi_0 - \gamma \sqrt{\phi_0}) \quad (7.7.7b)$$

†We remind the reader that the analysis presented assumed that the channel is in strong inversion. In the more general case, corresponding to the model of Sec. 4.3, we must allow for diffusion currents. Thus (7.7.6a) must be replaced by [see (4.3.15)]

$$q'_l(x, t) = -C'_{ox} \left[v_{GB}(t) - V_{FB} - \psi_s(x, t) - \gamma \sqrt{\psi_s(x, t)} \right]$$

where $\psi_s(x, t)$ is the surface potential at position x and at time t . Similarly, (7.7.6b) must be replaced by [see (4.3.5)]

$$i(x, t) = -\mu W q'_l(x, t) \frac{\partial \psi_s(x, t)}{\partial x} + \mu W \phi_t \frac{\partial q'_l(x, t)}{\partial x}$$

or, alternatively, (7.7.6b) can be used, with v_{CB} taken to be the “quasi-Fermi potential difference” (Appendix J). Equation (7.7.6c) is general and remains unchanged.

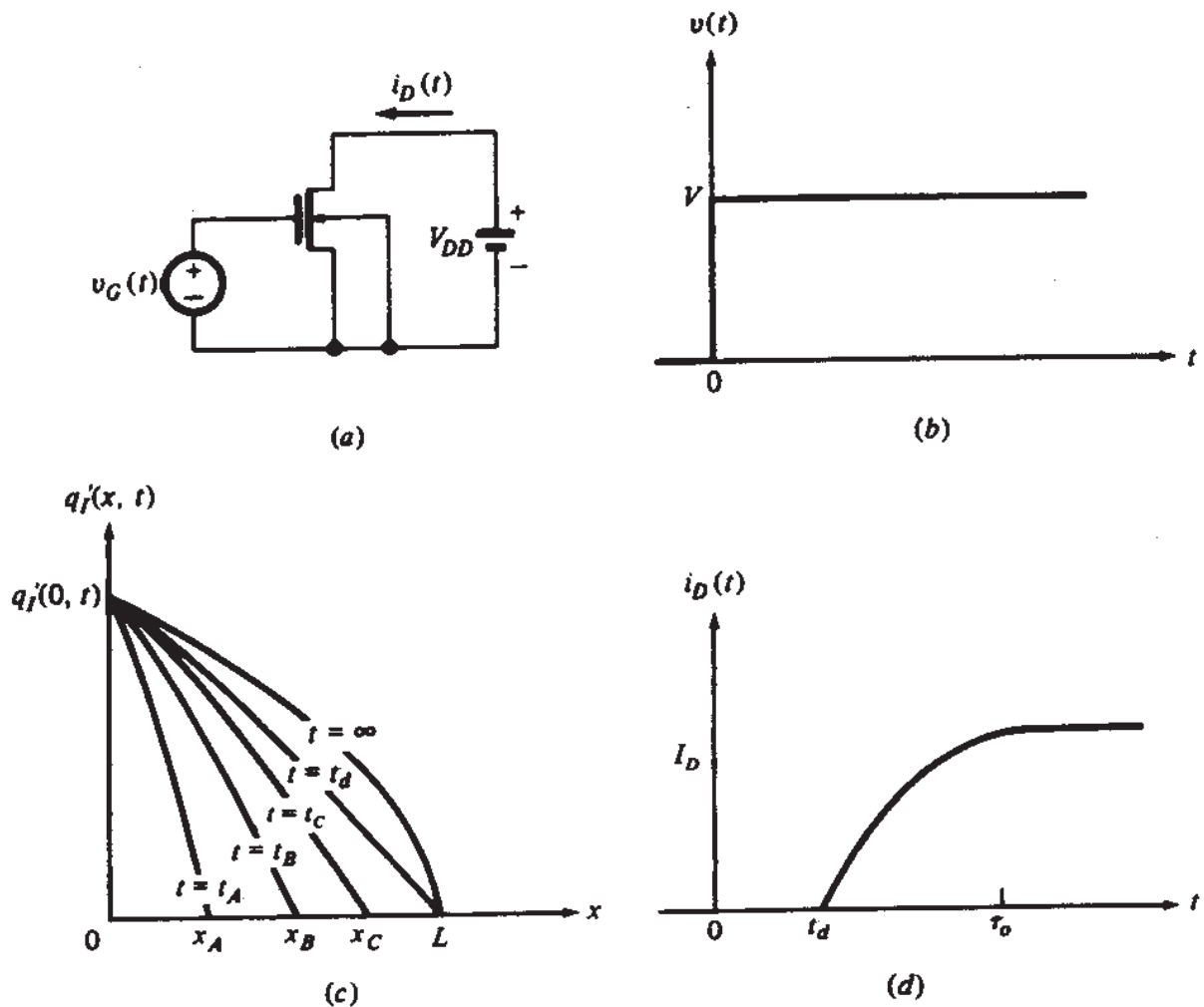


FIGURE 7.14

(a) A transistor with a step excitation; (b) gate-source voltage; (c) inversion layer charge per unit area as a function of position along the channel with time as a parameter; (d) drain current as a function of time. A sufficiently large V_{DD} is assumed, so that even the maximum value V of $v_G(t)$ corresponds to operation in the saturation region.

Finally, at the drain end ($x = L$), q'_I is zero whether the device is off or in saturation ("pinchoff" assumption). Thus

$$q'_I(L, t) = 0 \quad (7.7.7c)$$

The system (7.7.6) with the conditions (7.7.7) can now, in principle, be solved by using partial differential equation techniques to provide the distribution of i , v_{CB} , and q'_I with position and time. The drain and source currents can be determined from the solution for $i(x, t)$ by noting that

$$i_D(t) = i(L, t) \quad (7.7.8)$$

$$i_S(t) = -i(0, t) \quad (7.7.9)$$

From the solution for $v_{CB}(x, t)$ one can determine $q'_B(x, t)$ and integrate it with respect to position to find the total instantaneous depletion region charge as in (7.2.3c). The substrate transient current can then be determined as in (7.3.2). The gate

transient current can be found in a similar manner. We remind the reader that extrinsic parasitics are not considered here. In a real device, substrate and gate extrinsic resistance (Sec. 8.4) can change the picture.

Unfortunately, the actual details of the solution outlined above are complicated. The results presented in the literature are obtained by using numerical techniques. We will not present these long procedures here,^{6,12,19,20,22,32,54,62-69} but will only summarize the most important results. Solution of (7.7.6) for the circuit of Fig. 7.14a results in $q'_I(x, t)$, as shown in Fig. 7.14c. At $t = 0$, the channel is empty. At $t = t_A$, the electrons coming from the source have reached up to point $x = x_A$; hence, q'_I is zero beyond this point. The wavefront of electrons continues moving to the right as shown, and reaches the drain at $t = t_d$; t_d will be referred to as the *delay time*. At the instant $t = t_d$, the channel charge has not reached steady state yet. Steady state is reached asymptotically and corresponds to the curve marked $t = \infty$. Note that a quasi-static model implicitly assumes instead that this distribution is reached instantly at $t = 0^+$ ("an instant after" $t = 0$).

The drain current as a function of time is shown in Fig. 7.14d. It is zero up to $t = t_d$, at which time the electrons reach the drain. Current then begins to flow and, eventually, builds up to the value I_D , which is the value calculated from dc equations. [In contrast to the drain current, the source current starts flowing immediately at $t = 0^+$, since electrons start filling the channel from the source end as soon as $v(t)$ goes high, as expected by the fluid analog in Fig. 1.25d.]

Assuming q'_I corresponds to the approximate model of Sec. 4.5.3 with $\alpha = 1$, a numerical solution of the above problem gives for the delay time⁶

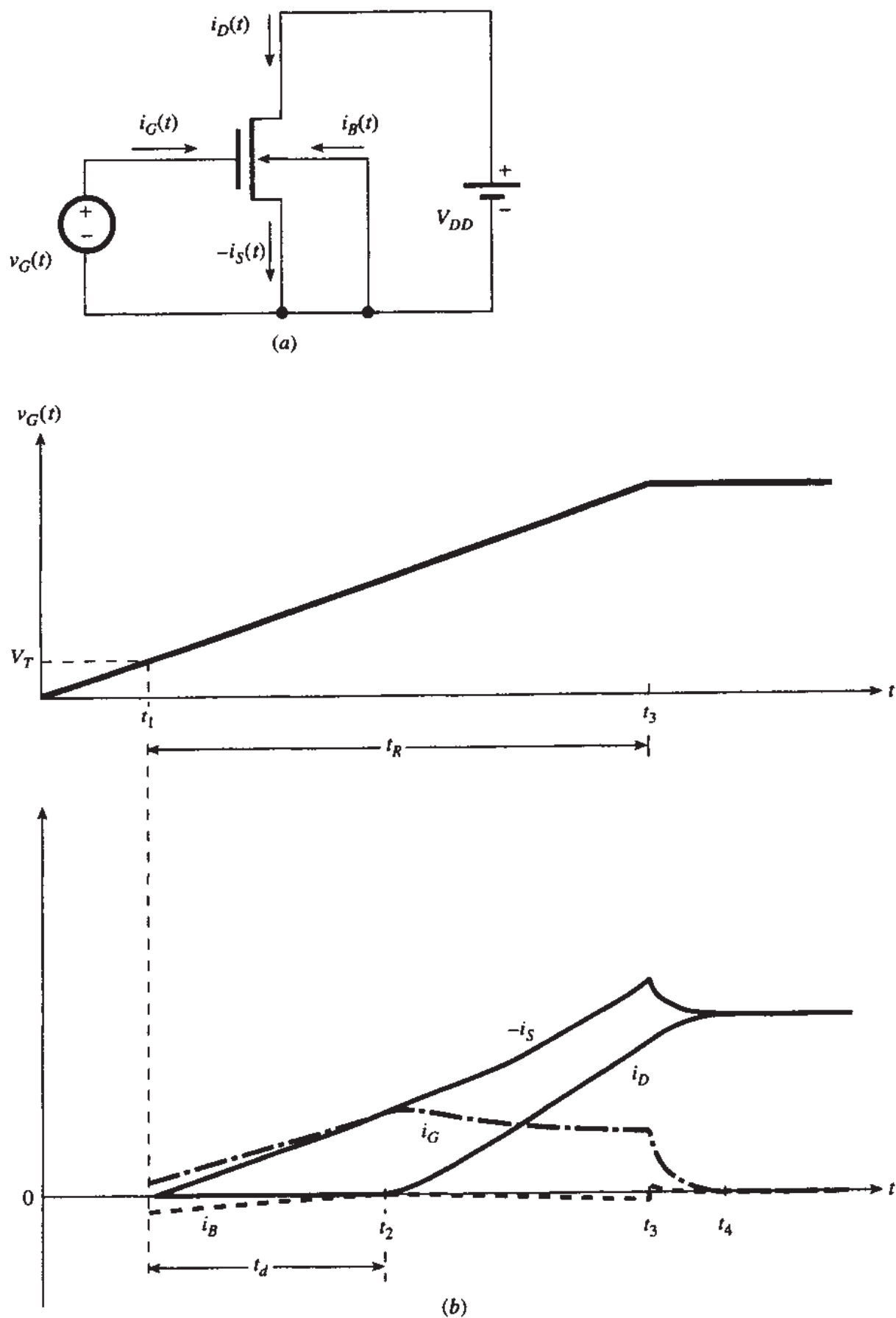
$$t_d \approx 0.38\tau_o \quad (7.7.10)$$

where

$$\tau_o = \frac{L^2}{\mu(V_{GS} - V_T)}, \quad \alpha = 1 \quad (7.7.11)$$

with V_{GS} being the value of the input for $t > 0$. The same numerical solution predicts that at $t = \tau_o$ the current has reached about 98 percent of the final value I_D . Note that the quantity τ_o in the above equation is the same as in (7.5.4), which was derived for dc conditions. However, the fact that the same quantity is involved in the two different types of analysis does not mean that one can casually use the dc transit time to explain directly the non-quasi-static behavior of the transistor. Care is required at this point.

We now briefly consider what happens when the above circuit is driven by an input with significant risetime t_R .^{12,19,20,22,32,54,64,66,68,69,71} In this simplified analysis, we will assume again that the transistor is off until a time $t = t_1$, when $v_{GS} < V_T$, and operates in strong inversion when $v_{GS} > V_T$. The value of V_{DD} is again assumed to be large enough to keep the device in saturation, even when the input attains its maximum value. The definitions of various currents are shown in Fig. 7.15a. In Fig. 7.15b, we show these currents vs. time for an input with t_R equal to several times τ_o . As seen, $-i_s$ (the current *leaving* the source) starts flowing as soon as the device turns on and is positive; positive current leaving the source is equivalent to negative current

**FIGURE 7.15**

(a) A transistor with a varying $v_G(t)$ and a fixed drain voltage (the latter is assumed to be large enough for the device to be in saturation, even with v_G at its maximum value); the currents plotted in the other parts of the figure are defined as shown here. (b) Waveforms for t_R equal to several times τ_o .

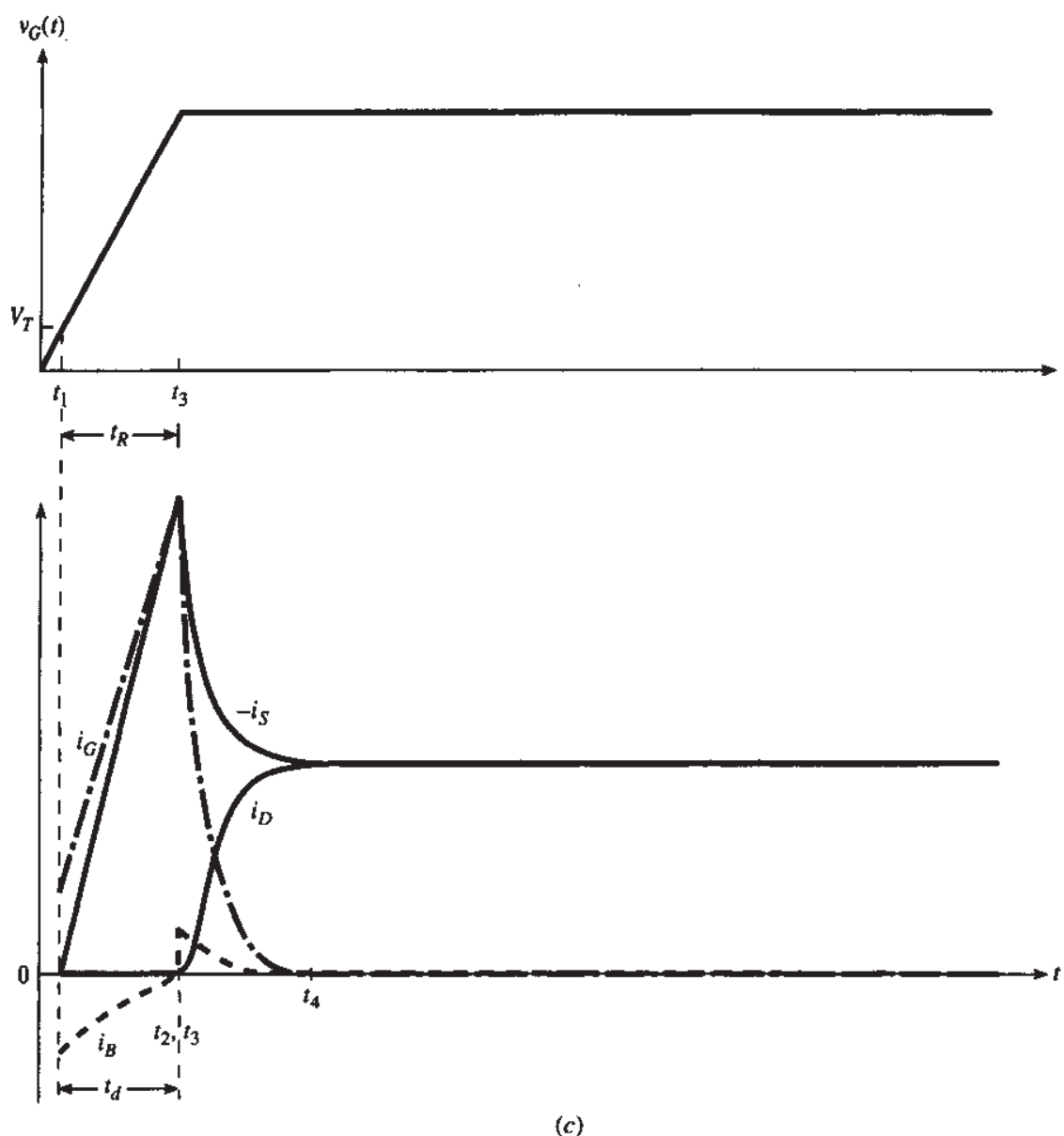


FIGURE 7.15—(Continued)

(c) Waveforms for t_R approximately equal to τ_o . Extrinsic capacitances are assumed zero, in order to clearly reveal the intrinsic transient behavior.

entering it; the channel of the device is been “charged” with mobile electrons. The current $-i_S$ soon rises to a value that is almost balanced by the gate current i_G ; the small difference seen is due to the substrate current, which is also shown in the figure. At time t_2 , the electrons reach the drain, and drain current begins to flow. This current approximately balances further increases in $-i_S$, and i_G does not increase any more. Since the channel must be charged further as v_G rises further, $-i_S$ continues to be larger than i_D . When the input stops rising at $t = t_3$, the channel is nearly charged, but not quite, since the inversion layer charge distribution has not reached steady state yet; thus $-i_S$ continues to be larger than i_D for a little while longer. However, $-i_S$ and i_D both gradually tend to their steady-state value, determined by the maximum

value of v_G using dc equations. At $t = t_4$ we have approximately $i_D = -i_S$ and $i_G = i_B = 0$; thus, we have practically reached dc steady state.

In Fig. 7.15c, the input rise time t_R has been reduced to a value of about τ_0 . The waveforms of the currents in this case should be compared to those in Fig. 7.15b. As seen, in Fig. 7.15c there is no time for i_G to taper off. Also, the larger value of dv_G/dt causes large charging currents (as would be the case even for a simple capacitor). Again, initially $-i_S$ rises to a value close to that of i_G , save for a small difference due to i_B . Once the maximum input value is reached, the behavior is qualitatively similar to that in Fig. 7.15b. Here, though, the mobile electrons happen to reach the drain at a time t_2 which is about equal to the time the input reaches its maximum value. Note that t_2 is smaller in Fig. 7.15c, since larger v_G values are available earlier than in Fig. 7.15b, and these help charge the channel faster.

As can be seen from the above example, the delay time t_d depends on the rise time t_R . As long as t_R is much less than τ_0 , the observations given earlier for the case of the step input are approximately valid. However, if t_R is significant (larger than τ_0), the delay time is given approximately by $t_d = \sqrt{t_R \tau_0}$, as found by another numerical solution. If t_R is over $20\tau_0$, the numerical solution of (7.7.6) gives roughly the same results as the quasi-static model. This is the reason why the limit of validity for the latter is expressed as in (7.6.3).

The above results have been derived for a long-channel device. If, instead, L is small, velocity saturation can occur, and electrons will be traveling at maximum speed $|v_d|_{\max}$ toward the drain. If it is assumed that this happens over the whole length of the channel, the delay time will be, for a step input,

$$\tau_d \approx \frac{L}{|v_d|_{\max}} \quad (7.7.12)$$

which can be significantly *larger* than what long-channel theory would predict.^{19,20} On the other hand, if the rise time of the input is significantly larger than the above limit, and saturation velocity is no longer the limiting factor, the delay is found to be less than that predicted by long-channel theory. One may attribute this to the fact that in short-channel devices the drain and source act also as gates, owing to two-dimensional effects (Sec. 6.3). These "gates" have already been activated before $t = 0$ in Fig. 7.14. Hence, the main gate does not have to start building up the whole inversion layer "from scratch." This heuristic explanation is supported by accurate numerical calculations.²⁰

REFERENCES

1. T. J. O'Reilly, "The transient response of insulated gate field-effect transistors," *Solid-State Electronics*, vol. 8, pp. 947-956, 1965.
2. Z. S. Gribnikov and Y. A. Tkhorik, "Calculation of the transient process in field triodes with an insulated gate for the saturated mode of operation," *Radio Engineering and Electronic Physics*, vol. 11, pp. 776-781, 1966.
3. A. Möschwitzer, "Zum statischen und dynamischen Grosssignalverhalten des MOS-Feldeffekt-Transistors," *NTZ*, vol. 20, pp. 150-154, 1967.

4. M. B. Das, "Switching characteristics of MOS and junction-gate field-effect transistors," *IEE Proceedings*, vol. 114, pp. 1223–1230, 1967.
5. F. A. Lindholm, R. J. Balda, and J. L. Clements, "Characterization of the four-terminal MOS transistor for digital and linear applications," *Digest of Technical Papers*, International Electronics Conference, Toronto, pp. 116–117, 1967.
6. J. R. Burns, "Large-signal transit-time effects in the MOS transistor," *RCA Review*, vol. 15, pp. 15–35, March 1969.
7. J. S. Brugler and P. G. A. Jespers, "Charge pumping in MOS devices," *IEEE Transactions on Electron Devices*, vol. ED-16, pp. 297–302, March 1969.
8. R. S. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley-Interscience, New York, 1970.
9. K. Gocer, "Einschaltzeiten und Umladungsvorgänge bei MOS-Transistoren," *AEU*, vol. 24, pp. 21–28, 1970.
10. D. J. Hamilton, F. A. Lindholm, and A. H. Marshak, *Principles and Applications of Semiconductor Device Modeling*, Holt, Rinehart, and Winston, New York, 1971.
11. J. E. Meyer, "MOS models and circuit simulation," *RCA Review*, vol. 32, pp. 42–43, March 1971.
12. M. E. Zahn, "Calculation of the turn-on behavior of MOST," *Solid-State Electronics*, vol. 17, pp. 843–854, 1974.
13. R. M. Swanson, "Complementary MOS transistors in micropower circuits," *Technical Report 4963-1*, Integrated Circuits Laboratory, Stanford University, California, 1974.
14. F. M. Klaassen, "A MOS model for computer-aided design," *Philips Research Reports*, vol. 31, pp. 71–83, 1976.
15. J. I. Arreola, "Equivalent circuit modeling of the large signal transient response of four-terminal MOS field-effect transistors," doctoral dissertation, University of Florida, 1978.
16. D. E. Ward and R. W. Dutton, "A charge-oriented model for MOS transistor capacitances," *IEEE Journal of Solid-State Circuits*, vol. SC-13, pp. 703–707, October 1978.
17. P. E. Cottrell and E. Buturla, "Two-dimensional static and transient simulation of mobile carrier transport in a semiconductor," *Proceedings of NASECODE I*, Dublin, June 1979.
18. J. A. Robinson, Y. A. El-Mansy, and A. R. Boothroyd, "A general four-terminal charging-current model for the insulated-gate field effect transistor," parts I and II, *Solid-State Electronics*, vol. 23, pp. 405–414, 1980.
19. S. Y. Oh, D. E. Ward, and R. W. Dutton, "Transient analysis of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. SC-15, pp. 636–643, August 1980.
20. S. Y. Oh, "A simplified two-dimensional numerical analysis of MOS devices including transient phenomena," *Technical Report G201-10*, Integrated Circuits Laboratory, Stanford University, California, June 1981.
21. D. E. Ward, "Charge-based modeling of capacitance in MOS transistors," *Technical Report G201-11*, Integrated Circuits Laboratory, Stanford University, California, June 1981.
22. M. S. Mock, "A time-dependent numerical model of the insulated-gate field-effect transistor," *Solid-State Electronics*, vol. 24, pp. 959–966, 1981.
23. Y. Ikawa, W. R. Eisenstadt, and R. W. Dutton, "Modeling of high-speed, large-signal transistor switching transients from s-parameter measurements," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 669–675, April 1982.
24. G. W. Taylor, W. Fichtner, and J. G. Simmons, "A description of MOS internodal capacitances for transient simulations," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. CAD-1, pp. 150–156, October 1982.
25. R. Conilogue and E. Viswanathan, "A complete large and small signal charge model for a M.O.S. Transistor," *Technical Digest*, International Electron Devices Meeting, San Francisco, pp. 654–657, 1982.
26. K. Y. Tong, "A model for MOS transistors from transient current computations," *IEE Proceedings*, vol. 130, part I, pp. 33–36, February 1983.
27. P. Yang, B. D. Epler, and P. K. Chatterjee, "An investigation of the charge conservation problem for MOSFET circuit simulation," *IEEE Journal of Solid-State Circuits*, vol. SC-18, pp. 128–138, February 1983.

28. C. Turchetti, G. Masetti, and Y. Tsividis, "On the small-signal behavior of the MOS transistor in quasi-static operation," *Solid-State Electronics*, vol. 26, pp. 941-949, 1983.
29. B. J. Sheu, D. L. Scharfetter, C. Hu, and D. O. Pederson, "A compact IGFET charge model," *IEEE Transactions on Circuits and Systems*, vol. CAS-31, pp. 745-749, August 1984.
30. J. J. Paulos, "Measurement and modeling of small-geometry MOS transistor capacitances," Ph.D. dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, September 1984.
31. C. Turchetti, P. Prioretti, G. Masetti, E. Profumo, and J. Vanzi, "A Meyer-like approach for the transient analysis of digital MOS IC's," *IEEE Transactions on CAD*, vol. CAD-5, pp. 499-507, October 1986.
32. C. Turchetti, P. Mancini, and G. Masetti, "A CAD-oriented non quasi-static approach for the transient analysis of MOS IC's," *IEEE Journal of Solid-State Circuits*, vol. SC-21, pp. 827-836, October 1986.
33. S. M. Sze, *Physics of Semiconductor Devices*, John Wiley & Sons, Inc., New York, 1981.
34. J. G. Fossum, H. Jeong, and S. Veeraraghavan, "Significance of the channel-charge partition in the transient MOSFET model," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 1621-1623, October 1986.
35. M. F. Sevat, "On the channel charge division in MOSFET modeling," *Digest of IEEE International Conference on Computer-Aided Design*, pp. 204-207, 1987.
36. M. A. Cirit, "The Meyer model revisited: why is charge not conserved?" *IEEE Transactions on Computer-Aided Design*, vol. CAD-8, pp. 1033-1037, 1989.
37. I. W. Smith, H. Statz, H. A. Haus, and R. A. Pucel, "On charge nonconservation in FET's," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 2565-2568.
38. K. A. Sakallah, Y.-T. Yen, and S. S. Greenberg, "A first-order charge conserving MOS capacitance model," *IEEE Transactions on Computer-Aided Design*, vol. 9, pp. 99-108, January 1990.
39. A. A. Kushaa and M. E. Nokali, "Modeling subthreshold capacitances of MOS transistors," *Solid-State Electronics*, vol. 35, pp. 45-49, 1992.
40. M. Anderson and P. Kuivalainen, "Transit-time model for short-channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. 40, p. 830, April 1993.
41. H. K. Lim and J. Fossum, "A charge-based large-signal model for thin-film SOI MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 446-457, February 1985.
42. J. Katzenelson and A. Unikovski, "A network charge-oriented MOS transistor model," *International Journal of High Speed Electronics and Systems*, vol. 6, pp. 285-316, 1995.
43. H. C. de Graaff and F. M. Klaassen, *Compact Transistor Modelling for Circuit Design*, Springer-Verlag, Vienna, 1990.
44. N. Arora, *MOSFET Models for VLSI Circuit Simulation—Theory and Practice*, Springer-Verlag, Vienna, 1993.
45. K. Lee, M. Shur, T. A. Fjeldly, and T. Ytterdal, *Semiconductor Device Modeling for VLSI*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
46. G. I. Serhan and S.-Y. Yu, "A simple charge-based model for MOS transistor capacitances: A new production tool," *IEEE Transactions on Computer-Aided Design*, vol. CAD-2, p. 48, January 1983.
47. M. A. Maher and C. A. Mead, "A physical charge-controlled model for MOS transistors," pp. 211-229 in P. Lesleben (editor), *Advanced Research in VLSI*, The MIT Press, Cambridge, Mass., 1987.
48. B. J. Sheu and P.-K. Ko, "Measurement and modeling of short-channel MOS transistor gate capacitances," *IEEE Journal of Solid-State Circuits*, vol. SC-22, pp. 464-472, June 1987.
49. H. Masuda, Y. Aoki, J. Mano, and O. Yamashiro, "MOSTSM: A physically based charge conservative MOSFET model," *IEEE Transactions on Computer-Aided Design*, vol. 7, pp. 1229-1236, December 1988.
50. B. J. Sheu, W.-J. Hsu, and P. K. Ko, "An MOS transistor charge model for VLSI design," *IEEE Transactions on Computer-Aided Design*, vol. 7, pp. 520-527, April 1988.

51. R. Gharabagi and M. E. Nokali, "A model for the intrinsic gate capacitances of short channel MOS-FETs," *Solid-State Electronics*, vol. 32, pp. 57–63, 1989.
52. W. Budde and W. H. Lamfried, "A charge-sheet capacitance model based on drain current modeling," *IEEE Transactions on Electron Devices*, vol. 37, pp. 1678–1687, July 1990.
53. R. Gharabagi, and M. A. E.-Nokali, "A charge-based model for short-channel MOS transistor capacitances," *IEEE Transactions on Electron Devices*, vol. 37, pp. 1064–1073, April 1990.
54. H.-J. Park, P. K. Ko, and C. Hu, "A charge conserving non-quasi-static (NQS) MOSFET model for SPICE transient analysis," *IEEE Transactions on Computer-Aided Design*, vol. 10, pp. 629–641, May 1991.
55. K.-M. Rho, K. Lee, M. Shur, and T. A. Fjeldly, "Unified quasi-static MOSFET capacitance model," *IEEE Transactions on Electron Devices*, vol. 40, pp. 131–135, January 1993.
56. D.-H. Cho, S.-M. Kang, K.-H. Kim, and S.-H. Lee, "An accurate intrinsic capacitance modeling for deep submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, vol. 42, pp. 540–548, March 1995.
57. A. I. A. Cunha, M. C. Schneider, and C. G.-Montoro, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electronics*, vol. 38, pp. 1945–1952, November 1995.
58. C. C. Enz, "The EVK model: A most model dedicated to low-current and low-voltage analogue circuit design and simulation," Chap. 7 in *Low-Power HF Microelectronics: A Unified Approach*, G. A. S. Machado (editor), IEE Circuits and Systems Series No. 8, IEE Book Publishing, 1996.
59. H.-H. Li, Y.-L. Chu, and C.-Y. Wu, "A new simplified charge pumping current model and its model parameter extraction," *IEEE Transactions on Electron Devices*, vol. 43, pp. 1857–1863, November 1996.
60. A. I. A. Cunha, O. C. Gouveia-Filho, M. C. Schneider, and C. Galup-Montoro, "A current-based model for the MOS transistor," *Proceedings 1997 International Symposium on Circuits and Systems*, pp. 1608–1611, Hong Kong, June 1997.
61. A. I. A. Cunha, *Um Modelo do Transistor MOS para Projecto de Cicuitos Integrados*, Ph.D. thesis, Universidade Federal de Santa Catarina, December 1996 (in Portuguese).
62. P. Mancini, C. Turchetti, and G. Masetti, "A non-quasi-static analysis of the transient behavior of the long-channel most valid in all regions of operation," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 325–334, February 1987.
63. K.-W. Chai and J. J. Paulos, "Unified nonquasi-static modeling of the long-channel four-terminal MOSFET for large- and small-signal analyses in all operating regimes," *IEEE Transactions on Electron Devices*, vol. 36, pp. 2513–2520, November 1989.
64. H. J. Park, P. K. Ko, and C. Hu, "A non-quasistatic MOSFET model for SPICE-Transient analysis," *IEEE Transactions on Electron Devices*, vol. 36, pp. 561–576, March 1989.
65. L.-J. Pu and Y. Tividis, "Harmonic distortion of the four-terminal MOSFET in non-quasistatic operation," *IEE Proceedings*, vol. 137, pp. 325–332, October 1990.
66. P. Roblin, S. C. Kang, and W.-R. Liou, "Improved small-signal equivalent circuit model and large-signal state equations for the MOSFET/MODFET wave equation," *IEEE Transactions on Electron Devices*, vol. 38, pp. 1706–1718, August 1991.
67. M. Chan, K. Hui, R. Neff, C. Hu, and P. K. Ko, "A relaxation time approach to model the non-quasi-static transient effects in MOSFET's," *Digest, International Electron Devices Meeting*, pp. 169–172, December 1994.
68. E. Dubois and E. Robilliart, "Efficient non-quasi-static MOSFET's model for circuit simulation," *Digest, International Electron Devices Meeting*, pp. 945–948, December, 1995.
69. W. Liu, C. Bowen, and M.-C. Chang, "A CAD-compatible non-quasi-static MOSFET model," *Digest, International Electron Meeting*, pp. 151–154, December 1996.
70. D. Foty, *MOSFET Modeling with Spice-Principles and Practice*, Prentice-Hall PTR, Upper Saddle River, N.J., 1997.
71. D. Murray, J. J. Sanchez, and T. A. Demassa, "A non-quasistatic semi-empirical model for small geometry MOSFETs," *Solid-State Electronics*, vol. 41, pp. 1309–1319, 1997.

PROBLEMS

- 7.1. Show that (7.4.4) and (7.4.5) are equivalent.
- 7.2. Derive in detail the expressions for the total charges Q_I , Q_B , Q_G , Q_D , and Q_S in (7.4.14), (7.4.15), (7.4.17), (7.4.19), and (7.4.20).
- 7.3. Consider a transistor with $N_A = 10^{17} \text{ cm}^{-3}$, $t_{ox} = 100 \text{ \AA}$, $Q'_o = 5 \times 10^{-9} \text{ C/cm}^2$, $\phi_{MS} = -1 \text{ V}$, $W = L = 2 \text{ }\mu\text{m}$, with $V_{SB} = 0 \text{ V}$. Plot the total charges Q_G , Q_B , Q_I , Q_D , and Q_S vs. V_{DS} , with V_{DS} between 0 and 5 V, for $V_{GS} = 3$ and 5 V. Use the results of Sec. 7.4.2.
- 7.4. Compare the empirical expressions for Q_D and Q_S in Appendix M to those corresponding to the approximate model [(7.4.19) and (7.4.20)].
- 7.5. Prove the expressions given for Q_B , Q_I , and Q_G corresponding to the complete strong-inversion model in Appendix M.
- 7.6. Compare the following three expressions for the total inversion layer charge Q_I :
 - (a) Empirical expression in Appendix M
 - (b) Expression corresponding to the simplified model, (7.4.14)
 - (c) Expression corresponding to the complete model in Appendix M
- 7.7. Prove the expressions for Q_I , Q_D , and Q_S in weak inversion given in Sec. 7.4.4.
- 7.8. Using the method outlined in Sec. 7.4.5, develop expressions for Q_I , Q_B , Q_G , Q_D , and Q_S in terms of ψ_{s0} and ψ_{sL} corresponding to the general model in Sec. 4.3.1 (this is a long problem).
- 7.9. Repeat Prob. 7.8 by using a first-degree polynomial approximation for Q'_B as in (4.3.25). Lengthy computations are needed.
- 7.10. For the device of Prob. 7.3, plot Q_G , Q_B , Q_D , and Q_S vs. V_{GS} , with V_{GS} from -2 to $+5 \text{ V}$, and for $V_{DS} = 1$ and 2 V .
- 7.11. Assume a fixed parasitic extrinsic capacitance C_{gde} exists between gate and drain in the device of Fig. 7.11a (such a capacitance can be caused by the physical overlap between gate and drain, as will be discussed in Sec. 8.4). Show that the effect of such a capacitance will be a downward shift of the current waveforms in Fig. 7.11d and e, between $t = 0$ and $t = t_3$.
- 7.12. Consider the transistor of Prob. 7.3, and assume a constant effective mobility $\mu = 600 \text{ cm}^2/(\text{V}\cdot\text{s})$. The transistor is connected as in Fig. 7.11a with $V_{DD} = 3 \text{ V}$; $v_G(t)$ is as shown in Fig. 7.11b, with $t_3 = 4 \text{ ns}$ and $v_{G,\max} = 3 \text{ V}$. Plot $i_D(t)$, $i_G(t)$, $i_S(t)$, and $i_B(t)$ as functions of time.

CHAPTER 8

SMALL-SIGNAL MODELING FOR LOW AND MEDIUM FREQUENCIES

8.1 INTRODUCTION

In the previous chapter we considered the MOS transistor with terminal voltages undergoing variations with time. No restrictions were placed on the magnitude of these variations. In this chapter we will consider the case where the terminal voltage variations are sufficiently small so that the resulting small current variations can be expressed in terms of them using linear relations. We will derive such linear relations and develop linear circuits to represent them. These circuits will be called *small-signal equivalent circuits*. When excited by voltages equal to the *small variations* of the actual terminal voltages, these circuits will produce currents equal to the *small variations* of the actual transistor currents. Such models find wide use in analog circuit design, and have been studied since the early days of MOS transistors.¹⁻⁵

In most of this chapter, we will concentrate on the *intrinsic* part of the transistor (Fig. 7.1). We will first develop a small-signal model valid when the voltage and current variations are so slow that charge storage effects can be neglected. Then we will develop a small-signal model valid at medium speeds, assuming quasi-static operation. Such models¹⁻⁴⁵ are sufficient for many applications. (More advanced models will be presented in Chap. 9.) It is convenient to discuss modeling by assuming that the voltage and current *variations* are sinusoidal. In that case, one often talks of

frequency rather than “speed,” and we will adopt this convention from now on. Following the discussion of the models for the intrinsic part of the transistor, we will consider modeling for the extrinsic part. We will then consider the noise generated within a MOS transistor.

As in the previous chapter, n -channel devices with long and wide channels on a uniform substrate and with constant mobility will be assumed, unless noted otherwise.

We begin by considering the *intrinsic* part of the device.

8.2 A LOW-FREQUENCY SMALL-SIGNAL MODEL FOR THE INTRINSIC PART

8.2.1 A Two-Path View

In developing a small-signal model for the dc drain current, we should recall that, in the general case, this current does not consist only of the channel (drain-to-source current) I_{DS} . As we have seen in Sec. 6.6, assuming $V_{DS} > 0$, high electric fields near the drain can result in impact ionization, which can cause a parasitic current I_{DB} to flow from the drain to the substrate. We assume that all of this current exits through the substrate terminal B . This current must be added to I_{DS} , in order to obtain the total drain current:

$$I_D = I_{DS} + I_{DB} \quad (8.2.1)$$

One can thus view the MOS transistor as containing two paths, as shown in Fig. 8.1. Since the channel path leads to the source, it is convenient, as we will see, to express I_{DS} as a function of voltages defined with respect to the source. Similarly, since the drain-to-substrate path leads to the substrate, it is convenient to express I_{DB} as a function of voltages defined with respect to the substrate. These conventions are indicated in the figure.†

Although the above approach could be generalized to include a gate current, as mentioned in Sec. 6.6 this current is negligible during normal operation. We thus assume for the dc gate current:

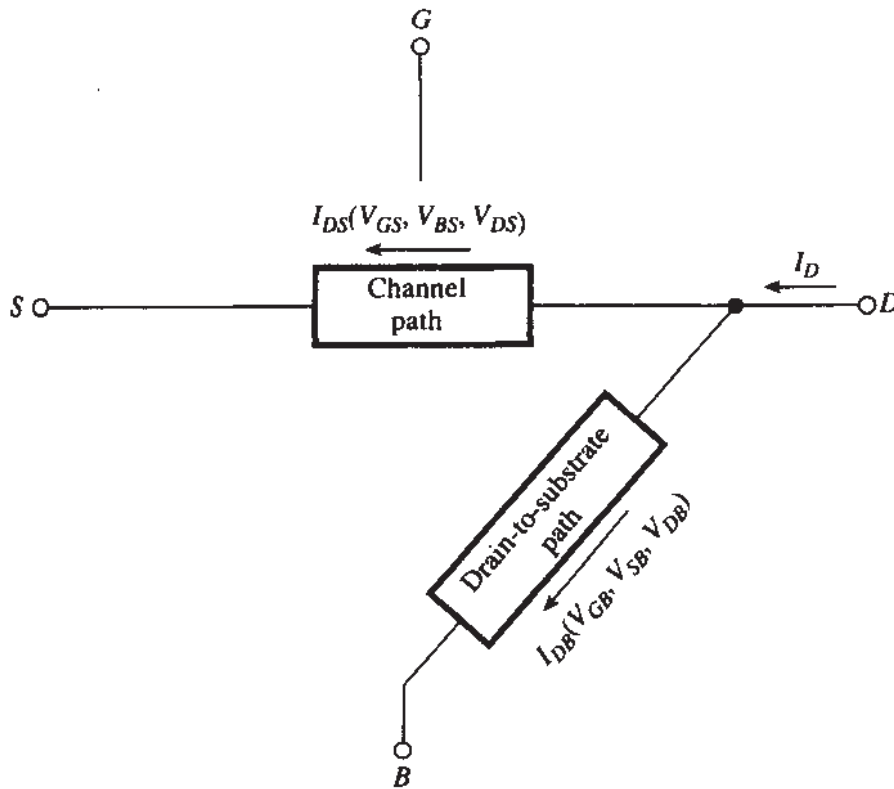
$$I_G = 0 \quad (8.2.2)$$

Assume that the terminal voltages are changed by a small amount. The resulting changes ΔI_D , ΔI_{DS} , and ΔI_{DB} in the dc values of the three currents are related by

$$\Delta I_D = \Delta I_{DS} + \Delta I_{DB} \quad (8.2.3)$$

From both this equation and from Fig. 8.1 it is clear that, for modeling purposes, we can consider each path separately, develop a small-signal model for each, and then combine the two models.

†We assume here that $V_{DS} > 0$, so that impact ionization occurs near the drain. For a symmetric model, one would need to include also a source-to-substrate path, which would be important if $V_{SD} > 0$.

**FIGURE 8.1**

The MOS transistor in dc operation, viewed as the combination of two paths.

8.2.2 A Small-Signal Model for the Channel Path

Let us consider an n -channel MOS transistor biased with V_{GS} , V_{SB} , and V_{DS} fixed at values V_{GS0} , V_{SB0} , and V_{DS0} , respectively, as shown in Fig. 8.2a (the subscript 0 indicates “quiescent” values). Let I_{DS0} be the resulting value of I_{DS} . We can study the effect of *very small* changes of the bias voltages on I_{DS} by varying these voltages *one at a time*, as shown in Fig. 8.2b, c, and d. We are for now interested only in the change of the *dc steady-state* value of I_{DS} ; that is, we assume that the voltages are constant before and after each change and that I_{DS} has reached dc steady state in both cases. We then consider the change ΔI_{DS} between the two dc steady-state values. We can relate cause and effect by using three conductance parameters, which can be measured as shown next to each figure. These parameters are

1. The *small-signal gate transconductance* g_m , often referred to simply as “transconductance.” Mathematically, it is defined by the following relation, corresponding to the situation in Fig. 8.2b:

$$g_m = \left. \frac{\partial I_{DS}}{\partial V_{GS}} \right|_{V_{BS}, V_{DS}} \quad (8.2.4)$$

where to the right of the vertical line we show the voltages being held fixed.[†]

[†]Strictly speaking, this “reminder” is redundant, since it is implied by the definition of a partial derivative, provided we remember that I_{DS} is expressed as a function of V_{GS} , V_{BS} , and V_{DS} . If, however, I_{DS} were expressed as a function of other independent variables (e.g., V_{GS} , V_{GD} , and V_{SB}), then $\partial I_{DS} / \partial V_{GS}$ would *not* mean the same as (8.2.4). So, for clarity, it doesn’t hurt to remind ourselves which voltages are being held constant.

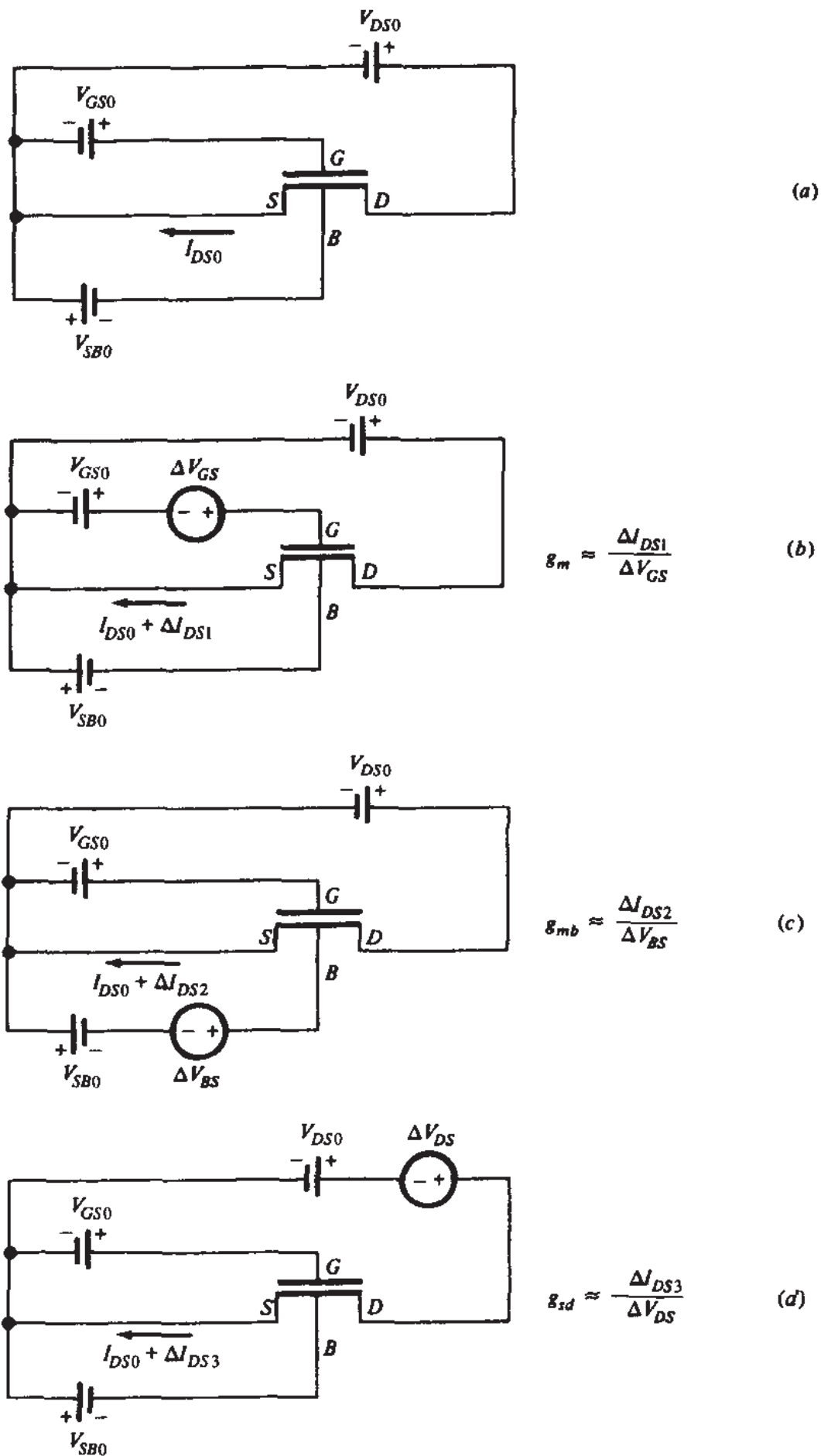
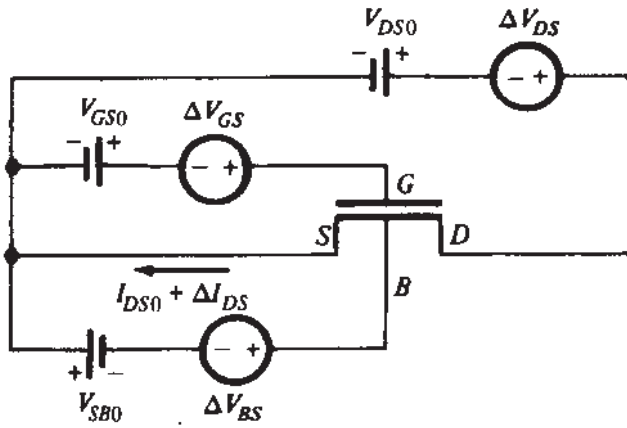


FIGURE 8.2

(a) A MOS transistor with dc voltages at a certain operating point. The rest of the figure shows the same circuit with very small changes (denoted by Δ) in the terminal voltages, and specifically (b) a small change in V_{GS} , (c) a small change in V_{BS} , (d) a small change in V_{DS} , and



(e)

FIGURE 8.2—(Continued)(e) small changes in V_{GS} , V_{BS} , and V_{DS} simultaneously.

2. The *small-signal substrate transconductance* g_{mb} . Corresponding to the situation in Fig. 8.2c, we have:

$$g_{mb} = \left. \frac{\partial I_{DS}}{\partial V_{BS}} \right|_{V_{GS}, V_{DS}} \quad (8.2.5)$$

Increasing V_{BS} as shown in Fig. 8.2c *decreases* V_{SB} . A consideration of the results of the body effect on I_{DS} shows that I_{DS} increases. Thus, ΔV_{BS} has on I_{DS} qualitatively the same effect as ΔV_{GS} has on Fig. 8.2b. The substrate acts, in this sense, as a second gate, and is often referred to as the *back gate*.

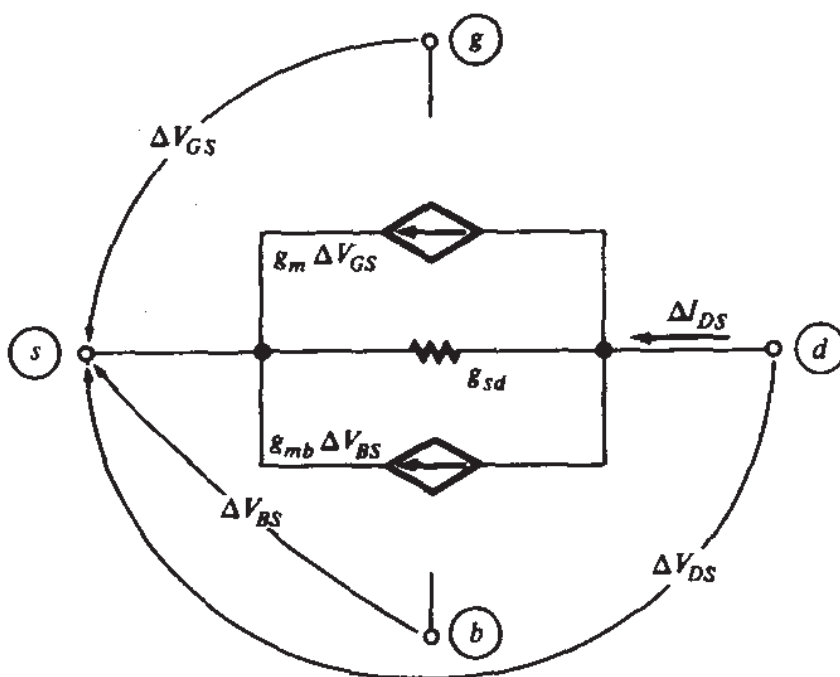
3. The *small-signal source-drain conductance* g_{sd} . Corresponding to the measurement in Fig. 8.1d, we define

$$g_{sd} = \left. \frac{\partial I_{DS}}{\partial V_{DS}} \right|_{V_{GS}, V_{BS}} \quad (8.2.6)$$

The approximate equality signs in Fig 8.2b to d become “equal” signs as ΔV_{GS} , ΔV_{BS} , and ΔV_{DS} approach zero. All three parameters have units of conductance and are expressed in S (S stands for Siemens; 1 S = 1 A/V).

Let us now consider the general case in which all three voltages are changed simultaneously, as shown in Fig. 8.2e. The corresponding total change in the drain-to-source current will be

$$\Delta I_{DS} \approx \left(\left. \frac{\partial I_{DS}}{\partial V_{GS}} \right|_{V_{BS}, V_{DS}} \right) \Delta V_{GS} + \left(\left. \frac{\partial I_{DS}}{\partial V_{BS}} \right|_{V_{GS}, V_{DS}} \right) \Delta V_{BS} + \left(\left. \frac{\partial I_{DS}}{\partial V_{DS}} \right|_{V_{GS}, V_{BS}} \right) \Delta V_{DS} \quad (8.2.7)$$

**FIGURE 8.3**

A low-frequency small-signal equivalent circuit for the *channel* of a MOS transistor. The drain-to-substrate path is not included in this partial model.

From our above definitions, this becomes

$$\Delta I_{DS} \approx g_m \Delta V_{GS} + g_{mb} \Delta V_{BS} + g_{sd} \Delta V_{DS} \quad (8.2.8)$$

In addition, (8.2.2) gives

$$\Delta I_G = 0 \quad (8.2.9)$$

The above two equations relate small-signal quantities and can be represented by the small-signal equivalent circuit in Fig. 8.3, where the rhombic symbols represent controlled current sources. It is easy to verify this circuit by writing Kirchhoff's current law for the right-hand node. As seen, we were able to represent the last term in (8.2.8) by a resistor of conductance g_{sd} , since ΔV_{DS} is the voltage *across* that resistor and thus $g_{sd} \Delta V_{DS}$ is the current through it. However, the other two terms in (8.2.8) have controlling voltages defined at nodes other than the node into which the current flows; thus, these terms are represented by voltage-controlled current sources.

In the preceding model derivation, the changes ΔV_{GS} , ΔV_{BS} , and ΔV_{DS} represented differences between two dc steady-state values of the terminal voltages V_{GS} , V_{BS} , and V_{DS} respectively. However, the model derived will be valid for representing the effects of gate and substrate on the drain current, even if the changes are continuously varying with time, as long as the variations are slow enough so that capacitive effects can be neglected. This will be understood better, and will be made more quantitative, after more complete models are considered. It will then be shown that, as the frequency of variation is decreased, more complete models reduce to the one in Fig. 8.3.

In the above definitions, we followed the common convention of referring the gate, drain, and substrate potentials to the source. An alternative approach, in which

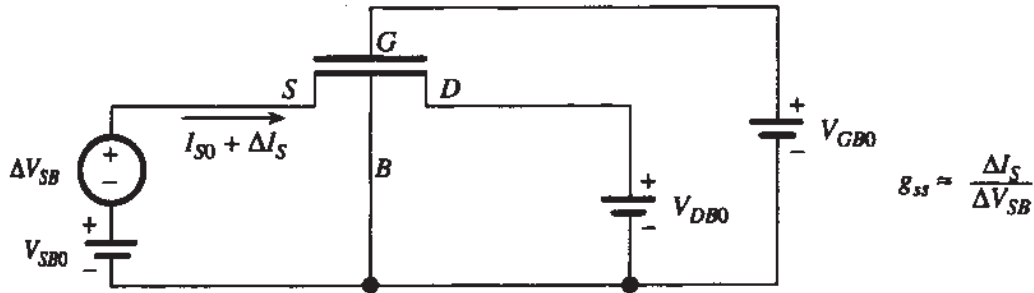


FIGURE 8.4
Measurement of source conductance.

all *four* terminal voltages are taken with respect to an arbitrary reference, is considered in Prob. 8.1.

In body-referenced models, a quantity used often,^{19,20} measured as shown in Fig. 8.4, is defined as follows:

$$g_{ss} = \left. \frac{\partial I_S}{\partial V_{SB}} \right|_{V_{GB}, V_{DB}} \quad (8.2.10)$$

where $I_S = -I_{DS}$. The quantity g_{ss} is seen to be the small-signal source conductance. Applying the small-signal equivalent circuit of Fig. 8.3 to the situation in Fig. 8.4, it is easy to show that (Prob. 8.3):

$$g_{ss} = g_m + g_{mb} + g_{sd} \quad (8.2.11)$$

8.2.3 A Small-Signal Model for the Drain-to-Substrate Path

One can follow an entirely analogous approach to model the drain-to-substrate path in Fig. 8.1. The role played above by the source will be played here by the substrate. With I_{DB} expressed as a function of V_{GB} , V_{SB} , and V_{DB} as indicated in that figure, we have, in analogy to (8.2.8)

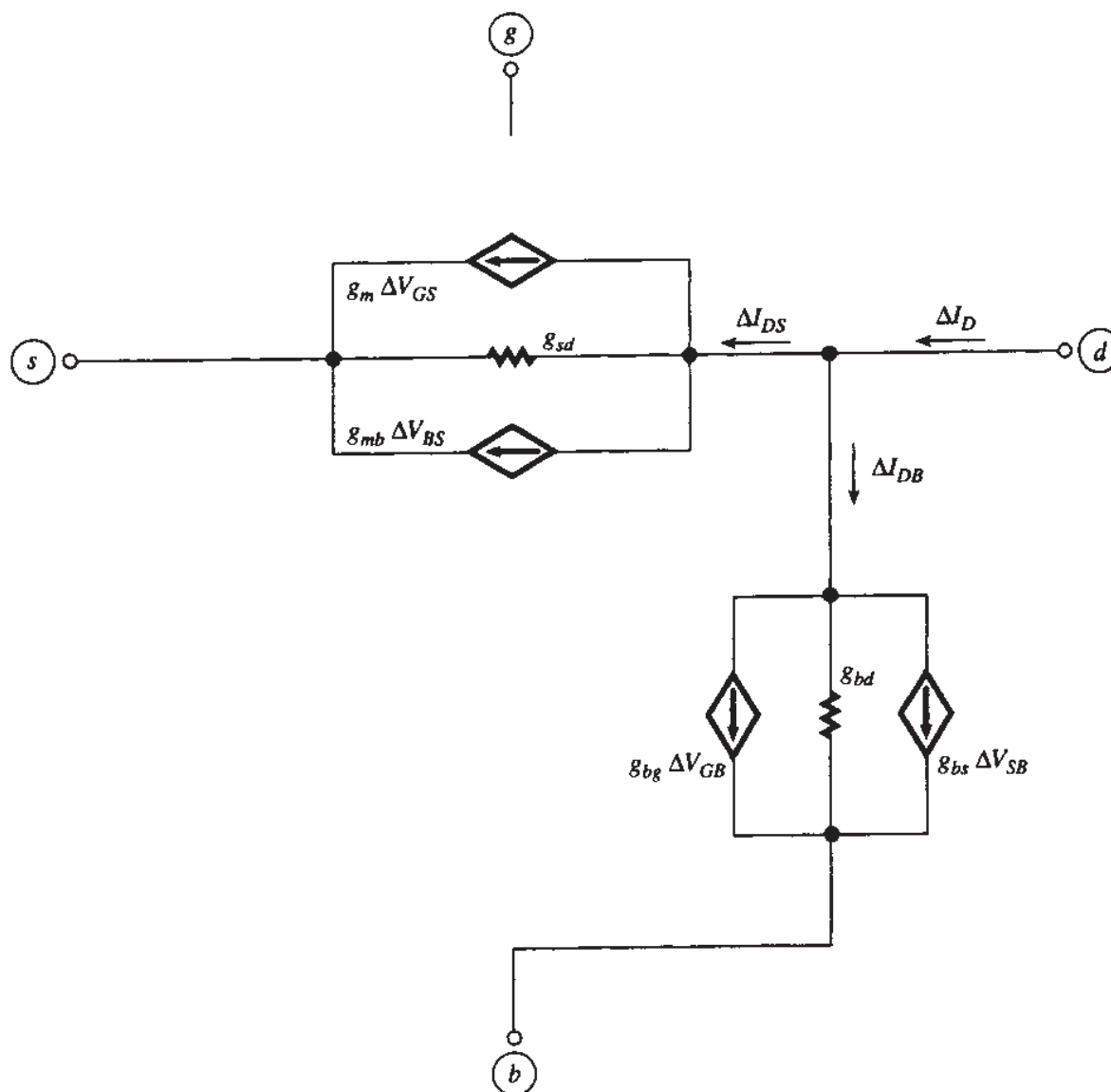
$$\Delta I_{DB} = g_{bg} \Delta V_{GB} + g_{bs} \Delta V_{SB} + g_{bd} \Delta V_{DB} \quad (8.2.12)$$

where

$$g_{bg} = \left. \frac{\partial I_{DB}}{\partial V_{GB}} \right|_{V_{SB}, V_{DB}} \quad (8.2.13)$$

$$g_{bs} = \left. \frac{\partial I_{DB}}{\partial V_{SB}} \right|_{V_{GB}, V_{DB}} \quad (8.2.14)$$

$$g_{bd} = \left. \frac{\partial I_{DB}}{\partial V_{DB}} \right|_{V_{GB}, V_{SB}} \quad (8.2.15)$$

**FIGURE 8.5**

A low-frequency small-signal model for the MOS transistor, including impact ionization effects.

Equation (8.2.12) can be modeled by three branches, in analogy to those used to model the channel above. Adding these three branches to the model of Fig. 8.3, we obtain the complete model of Fig. 8.5. It is easy to verify that the new elements model (8.2.12), and that the complete small-signal drain current ΔI_D satisfies (8.2.3). Although the I_{DB} -related elements are usually omitted from small-signal models, they can have important implications in analog circuit design (particularly the substrate-drain conductance g_{bd} , as will be seen).[†]

[†]One could have modeled the drain-to-substrate path by taking all voltages with respect to the source.⁴⁷ However, this would have resulted in three controlled sources. In the approach we have chosen, the element relating changes in V_{DB} to changes in I_{DB} is a two-terminal resistor, as seen in Fig. 8.5.

We will now derive expressions for the various small-signal parameters, considering each region of inversion separately. (Models valid in all regions of inversion will be considered in Sec. 8.2.7.) For brevity, the words *small signal* may be omitted when referring to a small-signal parameter when there is no chance for confusion, e.g., g_{sd} will be referred to as the *source-drain conductance*. Unless noted otherwise, in the derivations that follow no impact ionization will be assumed to be present.

8.2.4 Strong Inversion

GATE TRANSCONDUCTANCE. Consider a long-channel device with constant effective mobility. Using either the complete or the simplified model of Sec. 4.5 in the definition for g_m (8.2.4), we find

$$g_m = \frac{W}{L} \mu C'_{ox} V_{DS}, \quad V_{DS} \leq V'_{DS} \quad (8.2.16a)$$

$$= \frac{W}{L} \mu C'_{ox} V'_{DS}, \quad V_{DS} > V'_{DS} \quad (8.2.16b)$$

In the saturation region, the simplified model (4.5.37b) gives for the current $I_{DS} = \frac{1}{2}(W/L)\mu C'_{ox}(V_{GS} - V_T)^2/\alpha$. This, along with (8.2.4) produces the following equivalent forms† for g_m :

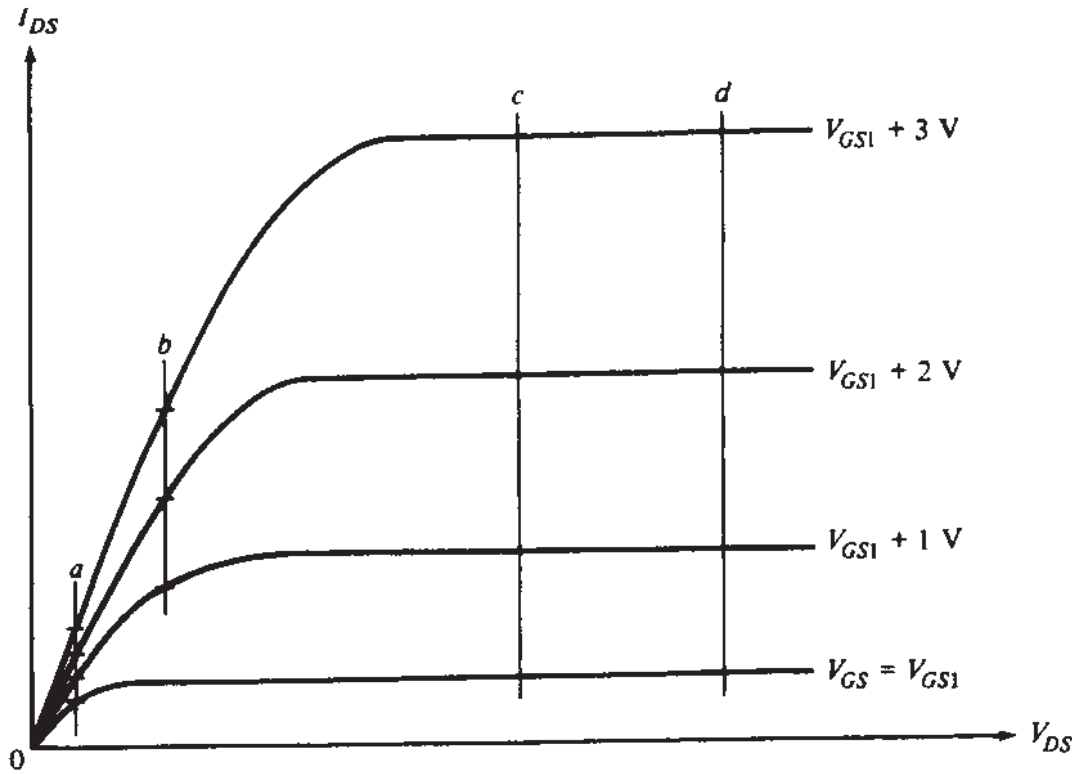
$$g_m = \frac{W}{L} \frac{\mu C'_{ox}}{\alpha} (V_{GS} - V_T), \quad V_{DS} > V'_{DS} \quad (8.2.17a)$$

$$= \sqrt{2 \frac{W}{L} \frac{\mu C'_{ox}}{\alpha} I_{DS}}, \quad V_{DS} > V'_{DS} \quad (8.2.17b)$$

$$= \frac{2I_{DS}}{V_{GS} - V_T}, \quad V_{DS} > V'_{DS} \quad (8.2.17c)$$

As noted from (8.2.16a), in nonsaturation g_m is independent of V_{GS} . This is illustrated graphically in Fig. 8.6. We assume that the V_{GS} step used to obtain successive curves was fixed, so that g_m can be estimated from the spacing of the curves if the V_{GS} steps are small. As seen, for lines *a* and *b* the spacing is independent of V_{GS}

†It is interesting to note that, at $V_{GS} = V_T$ and with I_{DS} calculated from (4.5.37b), these three equivalent expressions give g_m values of 0, $\sqrt{2(W/L)(\mu C'_{ox}/\alpha)I_{DS}}$, and infinity! This is a good example of the absurd results that can be obtained if modeling expressions are used outside their intended region of validity.

**FIGURE 8.6**

A family of I_{DS} - V_{DS} curves, obtained for fixed V_{GS} increment. In nonsaturation, spacing is proportional to V_{DS} but independent of V_{GS} (lines a , b); in saturation, spacing is independent of V_{DS} but depends on V_{GS} (lines c , d).

but depends on V_{DS} , as expected from (8.2.16a). In saturation, the situation is reversed, as seen from (8.2.16b) and illustrated by lines c and d . Now g_m is independent of V_{DS} but depends on V_{GS} (through V'_{DS}).

Various corrections can be applied to the above equations to increase their accuracy in the presence of higher order effects. For example, if the effective mobility is not constant with V_{GS} , differentiating I_{DS} to produce g_m will lead to an additional term (Prob. 8.6). An analogous correction should be applied if channel length modulation cannot be neglected. For example, a first-order correction results when obtaining g_m by differentiating (6.2.12).

Short-channel effects can strongly affect the value of all small-signal parameters. An extreme example is what happens to g_m in the presence of velocity saturation. The drain current is then roughly given by (6.5.14b). Differentiating that equation with respect to V_{GS} gives

$$g_m \approx WC'_{ox} \mu'_c, \quad \text{velocity saturation} \quad (8.2.18a)$$

$$\approx WC'_{ox} |v_d|_{\max}, \quad \text{velocity saturation} \quad (8.2.18b)$$

SUBSTRATE TRANSCONDUCTANCE. Using the definition of g_{mb} in (8.2.5) with the complete strong inversion model of Sec. 4.5.1, with $V_{DB} = V_{DS} + V_{SB}$, we get

$$g_{mb} = \left(\frac{\gamma}{\sqrt{V_{DS} + V_{SB} + \phi_0} + \sqrt{V_{SB} + \phi_0}} \right) g_m, \quad V_{DS} \leq V'_{DS} \quad (8.2.19a)$$

$$= \left(\frac{\gamma}{\sqrt{V'_{DS} + V_{SB} + \phi_0} + \sqrt{V_{SB} + \phi_0}} \right) g_m, \quad V_{DS} > V'_{DS} \quad (8.2.19b)$$

If V_{DS} is small, it can be neglected in (8.2.19a). Similarly, if V_{GS} is small, V'_{DS} will also be small, so V'_{DS} can be neglected in (8.2.19b). We thus obtain the following useful approximation:

$$\frac{g_{mb}}{g_m} \approx \frac{\gamma}{2\sqrt{V_{SB} + \phi_0}} = \frac{dV_T}{dV_{SB}} = \alpha_1 - 1 \approx n - 1, \quad \text{low } V_{DS} \text{ or } V_{GS} \quad (8.2.20a)$$

where, to obtain the last parts, we used (4.5.33), (4.5.41), and (4.6.16). Using the expression for γ from (3.4.3), as well as (2.2.4) and (3.4.8), it is easy to show that the above relation gives

$$\frac{g_{mb}}{g_m} \approx \frac{\epsilon_s}{\epsilon_{ox}} \frac{t_{ox}}{d_{Bm}}, \quad \text{low } V_{DS} \text{ or } V_{GS} \quad (8.2.20b)$$

where d_{Bm} is the depth of the depletion region, which is nearly uniform if V_{DS} is small (or if V_{GS} , and thus V'_{DS} , is small). The ratio g_{mb}/g_m can be thought of as a measure of the relative control of the “back gate” (substrate) and the “front gate.” As seen in (8.2.20b), the thinner the oxide the smaller the value of g_{mb}/g_m . This makes sense, since then the “front gate” is close to the channel and its control on it is strong. Thus the control of the “back gate” is smaller in a relative sense. The opposite is true if d_{Bm} is small. Then the edge of the depletion region is close to the channel and the substrate’s relative influence is large. In addition, the permittivities enter the picture. Large oxide permittivity means large control of the gate on the channel through the oxide, hence a small relative influence of the “back gate” (small g_{mb}/g_m). The opposite is true for large ϵ_s , which implies strong substrate control. The quantity g_{mb}/g_m is typically 0.1 to 0.3, but can rise significantly above those values for devices with large γ , operated with low V_{SB} .

It is not worth it to derive expressions for g_{mb}/g_m using the simplified model of (4.5.37), since the dependence of α on V_{SB} , discussed in Sec. 4.5.3, is *approximate*. It was developed to provide reasonable accuracy for I_{DS} , with no regard to the derivative of I_{DS} with respect to V_{BS} . “Reasonable accuracy” for a function does not necessarily imply reasonable accuracy for its derivatives. Such problems are common when differentiation of empirical or semiempirical expressions is attempted, and one must exercise caution. We will thus avoid obtaining g_{mb} by direct differentiation of the simplified drain current expression and rely only on (8.2.19).

If high-order corrections not involving V_{SB} are made on g_m [e.g., corresponding to (6.5.11)], then the corresponding corrections on g_{mb} will be taken care of

automatically through (8.2.19). Corrections involving V_{SB} [such as (4.10.20)] must, of course, be considered separately. In the presence of charge-sharing effects discussed in Sec. 6.3, approximate expressions for g_{mb} can be found by using approximate equations from that section in the definition (8.2.5). As might be expected, g_{mb}/g_m can become very small in the presence of strong short-channel effects, because in such cases the trapezoidal shape of the depletion region in Fig. 6.7 becomes close to a triangle, cutting off substrate control on the channel.

SOURCE-DRAIN CONDUCTANCE. The small-signal source-drain conductance g_{sd} ^{4,10-18,21-44} is the slope of the I_{DS} - V_{DS} characteristics with V_{GS} and V_{SB} held constant. Applying the definition (8.2.6) of g_{sd} to the nonsaturation complete model [(4.5.2b), with $V_{DB} = V_{DS} + V_{SB}$] gives

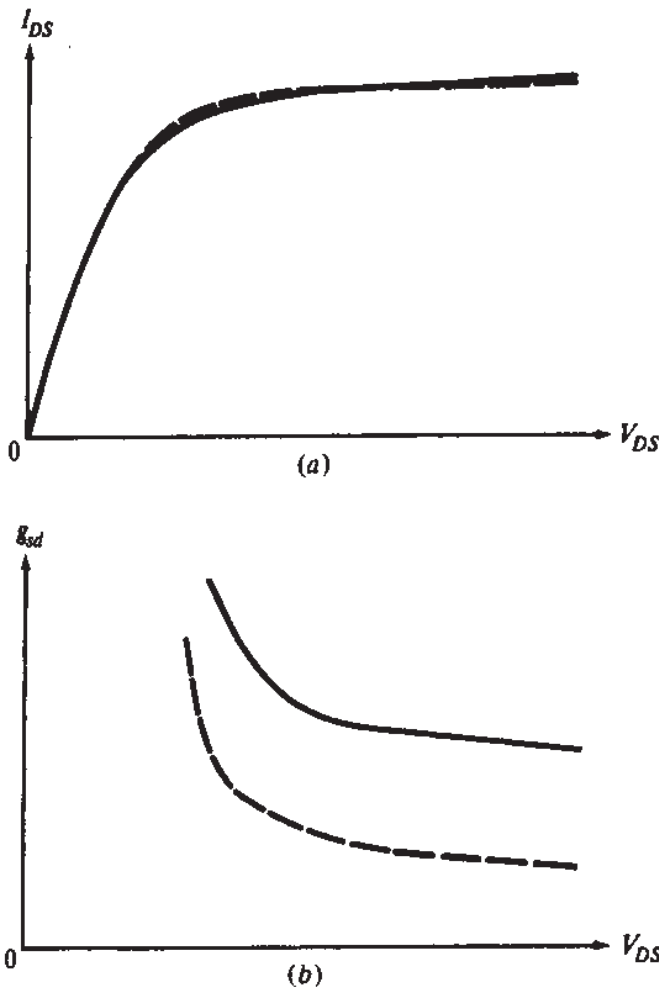
$$g_{sd} = \frac{W}{L} \mu C'_{ox} (V_{GS} - V_{DS} - V_{FB} - \phi_0 - \gamma \sqrt{V_{DS} + V_{SB} + \phi_0}), \quad V_{DS} \leq V'_{DS} \quad (8.2.21)$$

The simplified model (4.5.37a) gives

$$g_{sd} = \frac{W}{L} \mu C'_{ox} [V_{GS} - V_T - \alpha V_{DS}], \quad V_{DS} \leq V'_{DS} \quad (8.2.22)$$

The two expressions are equivalent for $V_{DS} = 0$.

In the saturation region, simple long-channel drain current models assume that I_{DS} is fixed at the value I'_{DS} (the value found from nonsaturation expressions with $V_{DS} = V'_{DS}$). The quantity I_{DS} is then predicted independent of V_{DS} , resulting in I_{DS} - V_{DS} curves parallel to the V_{DS} axis. Since g_{sd} is the slope of these curves, its value is predicted as zero, which is unacceptable. For many applications one needs instead a model that provides an accurate slope, which implies that the plot of I_{DS} vs. V_{DS} using such a model should track the nuances of experimentally observed behavior. This, unfortunately, represents a major problem area in MOS transistor modeling, as illustrated in Fig. 8.7.⁴⁵ In *a*, the solid line represents a measured characteristic. The broken line represents a model, the parameters of which have as usual been adjusted so that I_{DS} is predicted relatively accurately. Indeed, as shown, the error in predicting I_{DS} values is at most a few percent. In *b*, the *slopes* of the two curves are shown—the slope (g_{sd}) error exceeds 50 percent! This can have a very serious consequence in analog circuit design. For example, the small-signal “gain” of a CMOS inverter is inversely proportional to the sum of the g_{sd} ’s of two devices and can be in very large error if the above model is used to calculate it.

**FIGURE 8.7**

(a) I_{DS} - V_{DS} characteristics; solid line: measured; broken line: model; (b) slopes obtained from (a)⁴⁵ (© 1984 by IEEE).

If the saturation current is assumed to exhibit a first-order dependence on V_{DS} as in (6.2.12), the following g_{sd} value results:

$$g_{sd} \approx \frac{I'_{DS}}{V_A}, \quad V_{DS} > V'_{DS} \quad (8.2.23)$$

As expected from the above discussion, this formulation is too crude. In order to do a better job, we need to examine in more detail the phenomena responsible for the increase of I_{DS} with V_{DS} in saturation. In Chap. 6, we saw that two such phenomena are the channel length modulation (Sec. 6.2), and the drain-induced barrier lowering (Sec. 6.3). Although these phenomena appear simultaneously, we consider each one separately in order to provide some intuition through the use of simple relations.

To model channel length modulation (CLM), consider (6.2.5b), repeated here:^{4,12}

$$I_{DS} = \frac{I'_{DS}}{1 - l_p/L}, \quad V_{DS} > V'_{DS} \quad (8.2.24)$$

where l_p is the length of the "pinchoff" region. Using this in (8.2.6) we have (in the partial derivatives below, it is understood that V_{GS} and V_{BS} are kept constant)

$$\begin{aligned} g_{sd} &= \frac{\partial I_{DS}}{\partial V_{DS}} = \frac{\partial I_{DS}}{\partial l_p} \frac{\partial l_p}{\partial V_{DS}} \\ &= \frac{I'_{DS}}{(1 - l_p/L)^2} \frac{1}{L} \frac{\partial l_p}{\partial V_{DS}} \end{aligned} \quad (8.2.25)$$

or

$$g_{sd} = \frac{I_{DS}^2}{I'_{DS}} \frac{1}{L} \frac{\partial l_p}{\partial V_{DS}}, \quad V_{DS} > V'_{DS}, \quad \text{CLM only} \quad (8.2.26)$$

This, if I_{DS}/I'_{DS} is not very different from unity, simplifies to

$$g_{sd} \approx I'_{DS} \frac{1}{L} \frac{\partial l_p}{\partial V_{DS}}, \quad V_{DS} > V'_{DS}, \quad \text{CLM only} \quad (8.2.27)$$

Various expressions for g_{sd} can be found, depending on the behavior assumed for l_p (Sec. 6.2). For example, if (6.2.6) is used, we obtain

$$\frac{\partial l_p}{\partial V_{DS}} = \frac{B_1}{\sqrt{N_A}} \frac{1}{2\sqrt{\phi_D + V_{DS} - V'_{DS}}} \quad (8.2.28)$$

Thus, using (8.2.27), we obtain

$$g_{sd} = \frac{B_1 I'_{DS}}{2L\sqrt{N_A}\sqrt{\phi_D + V_{DS} - V'_{DS}}} \quad (8.2.29)$$

This level of modeling is widely used in computer-aided MOS circuit analysis and design, and is sometimes adequate for digital circuits. For analog circuit applications, though, it is still not satisfactory, because such applications can place heavy demands on the models of g_{sd} in saturation, as already explained. The reason that this parameter resists attempts to model it as above is that near the drain the electric field distribution is actually *two-dimensional*. The gradual channel approximation fails in that vicinity, and the field can have a significant horizontal component. The details of the field there are influenced by the drain region details (*junction depth*, for example) and, notably, by *field lines emanating from the gate*. The result is a current flow which, at least in part, takes place in a subsurface path. The associated charge in the "pinchoff" region should be taken into account for very accurate g_{sd} modeling. For an accurate evaluation of I_{DS} (and thus g_{sd}), then, one must resort to *two-dimensional*

numerical simulations using a computer.¹⁰ Pseudo-two-dimensional approaches have also been proposed.^{12,17,18,21,23,25,26,42,43} One such analysis^{42,43} leads to the result for l_p , given in (6.2.11). Using this in (8.2.27) we obtain:

$$g_{sd} = \frac{l_a}{L} \frac{I'_{DS}}{V_E + (V_{DS} - V'_{DS})} \quad (8.2.30)$$

where l_a depends on junction depth, as discussed in Sec. 6.2, and V_E is best determined from measurements. Typically, V_E is less than 1 V. Note that the above equation can be written in the form of (8.2.23) as follows:

$$g_{sd} = \frac{I'_{DS}}{V_A(V_{DS})} \quad (8.2.31a)$$

where

$$V_A(V_{DS}) = \frac{L}{l_a} [V_E + (V_{DS} - V'_{DS})] \quad (8.2.31b)$$

Let us now consider the effect of drain-induced barrier lowering (DIBL, Sec. 6.3) on g_{sd} . In order to reveal the main effects of this phenomenon, we will neglect channel length modulation in the following discussion. We will assume that no velocity saturation occurs, and that the device can be described in saturation by the simple square-law equation

$$I_{DS} = \frac{W}{L} \frac{\mu C'_{ox}}{2\alpha} [V_{GS} - \hat{V}_T(V_{DS})]^2, \quad V_{DS} > V'_{DS} \quad (8.2.32)$$

where $\hat{V}_T(V_{DS})$ was seen in Sec. 6.3 to depend on V_{DS} , and was modeled using charge sharing or DIBL concepts. It was seen that \hat{V}_T decreases as V_{DS} is raised, even in the saturation region. This gives a positive slope to the I_{DS} vs. V_{DS} plot, i.e., a positive g_{sd} . Differentiating (8.2.32) with respect to V_{DS} we obtain

$$g_{sd} = \frac{W}{L} \frac{\mu C'_{ox}}{\alpha} [V_{GS} - \hat{V}_T(V_{DS})] \left(-\frac{\partial \hat{V}_T}{\partial V_{DS}} \right) \quad (8.2.33)$$

We will relate this result to the gate transconductance g_m . Using (8.2.32) in (8.2.4) we find that g_m is given by (8.2.17a), with V_T replaced by $\hat{V}_T(V_{DS})$. From this and (8.2.33) we obtain

$$g_{sd} = g_m \left(-\frac{\partial \hat{V}_T}{\partial V_{DS}} \right), \quad V_{DS} > V'_{DS}, \quad \text{DIBL only}$$

(8.2.34)

To obtain a qualitative feel, let us use the simple model of (6.3.3) and (6.3.15) for $\hat{V}_T(V_{DS})$. This gives

$$\frac{g_{sd}}{g_m} \approx 0.5 \frac{\epsilon_s}{\epsilon_{ox}} \frac{t_{ox}}{L} \quad (8.2.35)$$

Qualitatively, the behavior predicted by this equation can be reached by using a different picture,³ which assumes that some field lines from the drain curve down, then bend and turn up, and terminate directly on inversion layer charges throughout much of the channel. Thus, the drain acts as a “somewhat inefficient gate”; increasing the voltage on this “gate” increases the current and gives rise to a nonzero g_{sd} . The parameter g_{sd}/g_m is a measure of the “competition” between this “gate” and the normal gate (just like g_{mb}/g_m played a similar role as a measure of the relative competition between the substrate and the normal gate above). The smaller the L , the closer the whole of the inversion layer is to the drain region, and the stronger the influence of the latter; g_{sd}/g_m can be large in this case. The smaller the t_{ox} , the closer the actual gate is to the inversion layer, and the more this gate wins out in the competition; then g_{sd}/g_m will be small. The relative influence of the two gates also depends on the permittivity of the media separating them from the inversion layer: Large permittivity implies a large effective capacitance and, hence, a strong influence. All these qualitative results are consistent with (8.2.35). Similar qualitative conclusions can be reached if the model of (6.3.17) and (6.3.18) is used for $\hat{V}_T(V_{DS})$, which gives:

$$\frac{g_{sd}}{g_m} \approx \exp \left[-\sqrt{\frac{\epsilon_{ox}}{\epsilon_s} \frac{L^2}{t_{ox} d_B}} \beta_3 \right] \quad (8.2.36)$$

Finally, note that the larger the drain junction depth, the stronger will be the drain’s “presence” as far as the inversion layer is concerned. Hence, g_{sd} , and g_{sd}/g_m , should *increase* with increasing drain junction depth. Dependence on junction depth is predicted by “charge-sharing” models such as those in Sec. 6.3.2, provided junction depth is retained as a parameter (Prob. 8.11).

Although the effects of channel-length modulation and DIBL were discussed separately above, in a real device they are of course present simultaneously. In general, DIBL can be dominant for V_{GS} close to V_T , with channel length modulation becoming dominant at higher V_{GS} values. For a complete description of g_{sd} , one would have to differentiate a complete expression for I_{DS} , including all short-channel effects. This can lead to very complicated expressions involving all device geometrical parameters, substrate doping, saturation velocity, and even inversion layer thickness.¹⁷ The resulting expression is very much dependent on the form of the model used to derive it.

The transition from nonsaturation to saturation is especially difficult to model (i.e., for values of V_{DS} close to V'_{DS}). As has been remarked in Sec. 4.7, strong-inversion models are not accurate in that region, since the drain end of the channel is then moderately inverted. Thus, if strong-inversion expressions are used in that region, errors in g_{sd} can result.

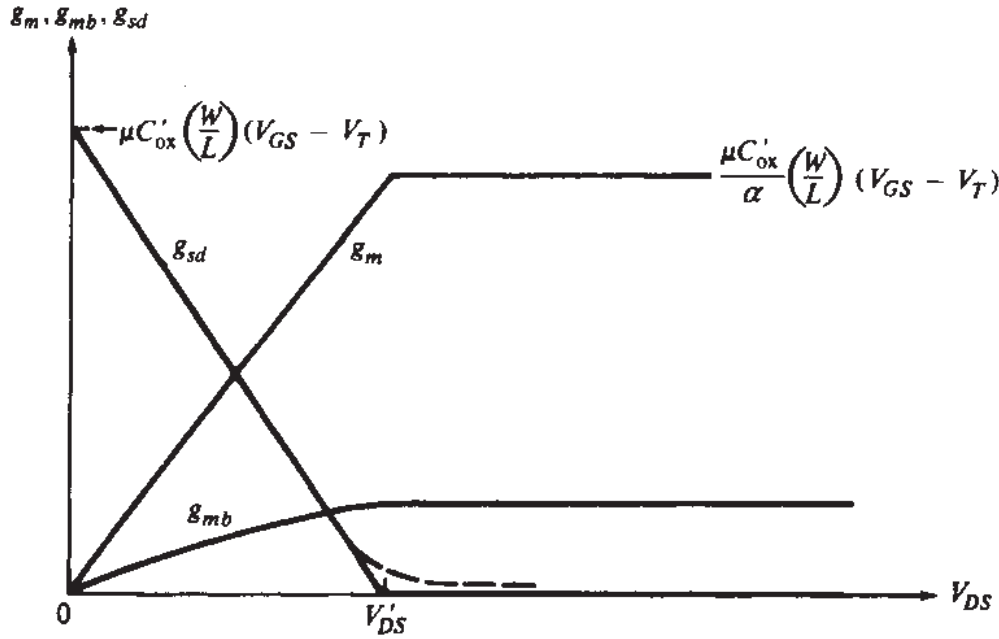


FIGURE 8.8

Small-signal gate transconductance, substrate transconductance, and drain conductance versus V_{DS} for fixed V_{GS} and V_{SB} , as predicted by the simplified model (solid lines). The broken line represents a more accurate modeling for g_{sd} .

It should be emphasized that, if separate I_{DS} expressions are used to derive g_{sd} for different regions, it should be made sure that the resulting expressions for g_{sd} are continuous at the region boundaries, and that at least the first derivatives of g_{sd} with respect to each terminal voltage are also continuous (see also Chap. 10). In some computer simulator models, “smoothing functions” are used to produce a single I_{DS} expression out of the separate expressions for each region.⁴⁶

Figure 8.8 shows the small-signal conductance parameters g_m , g_{mb} , and g_{sd} plotted vs. V_{DS} for a fixed V_{GS} , using simple nonsaturation expressions derived above up to $V_{DS} = V'_{DS}$, and assuming that the parameters are constant above that point. The sharp corners are, of course, artificial and are the result of the simplifications involved in the modeling process. More accurate models would provide smooth curves. For example, if one models correctly the transition region between nonsaturation and saturation, and includes channel length modulation in saturation, the g_{sd} plot would be of the form shown by the dashed line in the figure.

It is interesting to note that, as follows from Fig. 8.8,

$$g_{sd}|_{V_{DS}=0} \approx g_m|_{V_{DS} > V'_{DS}}, \quad \text{if } \alpha \approx 1 \quad (8.2.37)$$

assuming, of course, the same V_{GS} value.

SUBSTRATE-DRAIN PATH ELEMENTS. As already discussed, the effect of impact ionization can be modeled by using the three elements in the lower part of Fig. 8.5. The values of these can be obtained by using the definitions (8.2.13) to (8.2.15) with appropriate expressions for I_{DB} (Sec. 6.6). From (8.2.13) and Fig. 6.26 it is seen that

g_{bg} will be positive for V_{GS} values below the peak of I_{DB} , and negative for values above that. The magnitude of g_{bg} in normal operation is much smaller than g_m , and its effect is usually negligible. Similarly, g_{bs} is usually negligible. However, the effect of g_{bd} can be very important, as we will see shortly. Using (6.6.1) in (8.2.15), and keeping the dominant terms, we obtain:

$$g_{bd} = \frac{I_{DB} V_i}{(V_{DS} - V'_{DS})^2} \quad (8.2.38)$$

where V_i is typically 10 to 30 V, and I_{DB} is normally several orders of magnitude below I_D . This formula is very approximate, and should be used only for rough calculations.

OUTPUT CONDUCTANCE. Consider a transistor connected as shown in Fig. 8.9a. Its total drain current, I_D , is shown versus V_{DS} in Fig. 8.9b. Let us define the slope of this plot as the small-signal "output conductance" g_o :

$$g_o = \left. \frac{\partial I_D}{\partial V_{DS}} \right|_{V_{GS}, V_{SB}} \quad (8.2.39)$$

This quantity is shown plotted versus V_{DS} in Fig. 8.9c. The main effects responsible for the behavior of this plot are marked on the figure. There are two ways in which impact ionization can influence the value of g_o and make it rise with V_{DS} . One is directly through the change in I_{DB} . Limiting attention to the intrinsic device, it is easy to see that $g_o = g_{sd} + g_{bd}$. This can be seen from the small-signal equivalent circuit in Fig. 8.5, by using in it $\Delta V_{GS} = \Delta V_{GB} = \Delta V_{SB} = \Delta V_{BS} = 0$, and connecting the substrate to the source, corresponding to Fig. 8.9a.

A second effect⁴⁷ is present due to the resistance of the substrate material (this resistance belongs to the "extrinsic" part of the device, which will be discussed in Sec. 8.4; however, it is appropriate to mention its effect here). Assume that the substrate offers a resistance R_{be} to the substrate current I_{DB} . This causes a voltage drop $R_{be} I_{DB}$, which opposes the source-substrate voltage V_{SB} ; i.e., the intrinsic transistor now does not see V_{SB} , but rather an effective voltage $V_{SB} - R_{be} I_{DB}$. As an increasing drain voltage increases I_{DB} , the above effective voltage decreases, which decreases the threshold voltage due to the body effect. This then causes an extra increase in I_{DS} , not accounted for by the effects considered so far. One easily finds that, assuming $R_{be} \ll 1/g_{bd}$, the output conductance g_o is given by (Prob. 8.12):

$$g_o \approx g_{sd} + g_{mb} R_{be} g_{bd} + g_{bd} \quad (8.2.40)$$

This quantity is difficult to calculate accurately, as the path over which I_{DB} flows changes with bias, thus changing the effective R_{be} . For devices built on a low-resistivity epitaxial layer, R_{be} is small and so is the middle term in the right-hand side of the above equation.

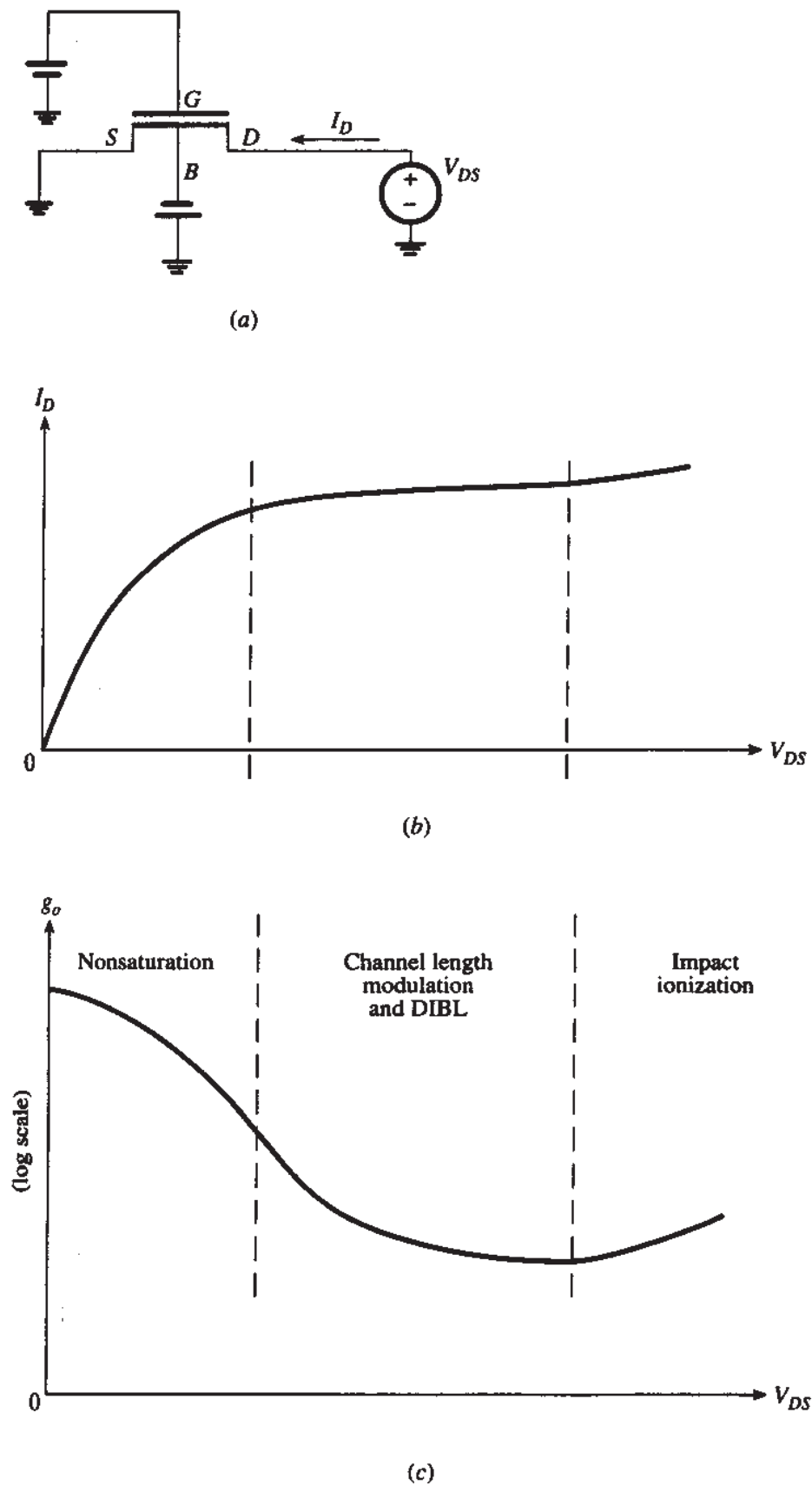
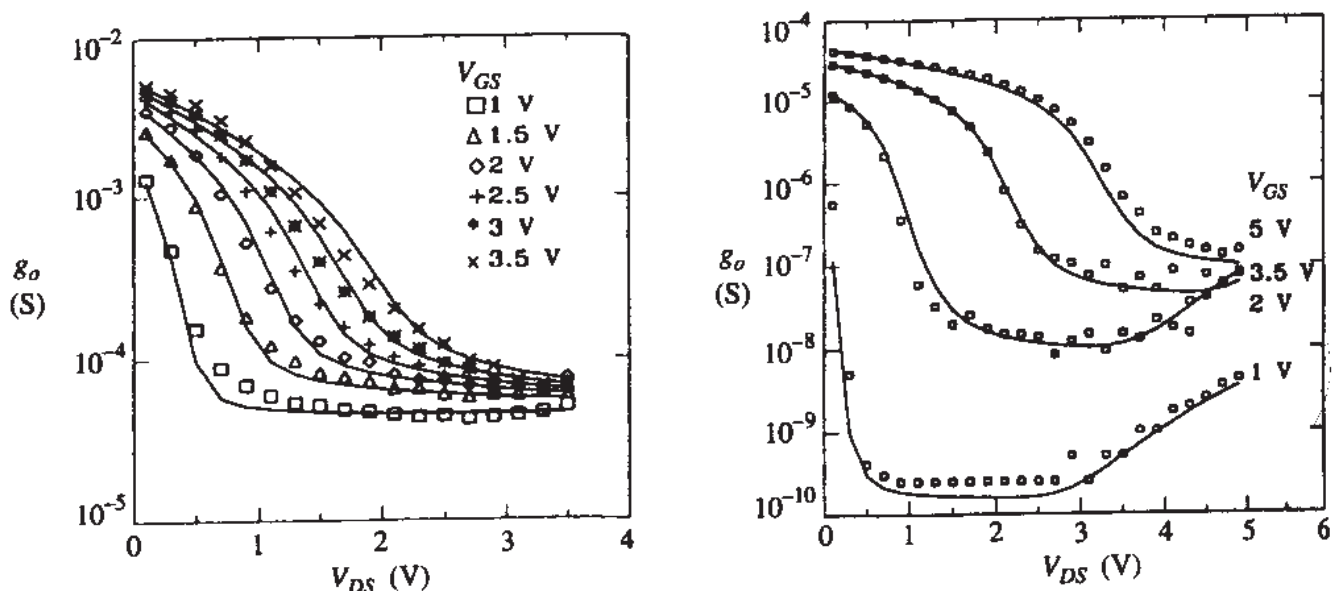


FIGURE 8.9

(a) Transistor with fixed V_{GS} and V_{SB} , and a variable V_{DS} . (b) Total drain current I_D versus V_{DS} for the connection in (a). (c) Output conductance [slope of the plot in (b)], with dominant effects marked.

**FIGURE 8.10**

Output conductance (on a log scale) (a) for a short-channel device ($W = 10 \mu\text{m}$, $L = 0.35 \mu\text{m}$, $t_{\text{ox}} = 100 \text{ \AA}$) and (b) for a long-channel device ($W = 1.2 \mu\text{m}$, $L = 10 \mu\text{m}$, $t_{\text{ox}} = 150 \text{ \AA}$)³⁵ (© 1993 by IEEE).

The quantity g_o is shown versus V_{DS} , with V_{GS} as a parameter, in Fig. 8.10, for a short- and long-channel device.³⁵ For the former, the presence of strong channel-length modulation and DIBL effects keeps g_{sd} large, no matter what bias values are used. In contrast, the long-channel device g_{sd} can vary over a very wide range and can be made very small for V_{DS} of a couple of volts, in fact smaller than the other terms in (8.2.40). This helps reveal those terms and is the reason that impact ionization effects are more apparent in this device. As seen, impact ionization can increase g_o by over an order of magnitude. This can have very serious consequences in analog circuit design, so it is very important to include g_{bd} in the small-signal equivalent circuit of the intrinsic device (and, of course, R_{be} in the model of the complete device—see Sec. 8.4).

APPARENT TRANSCONDUCTANCE “OVERSHOOT.” The effect of the substrate extrinsic resistance, referred to above, does not only affect g_o . When V_{GS} is increased, I_{DB} initially increases, as shown in Fig. 6.26. This decreases the effective source-substrate bias $V_{SB} - R_{be}I_{DB}$, and thus decreases the effective threshold, causing an excess increase in I_{DS} , above what would be expected in the absence of R_{be} . If the origin of this effect is not known, the increase in $\partial I_D / \partial V_{GS}$ might be interpreted as a “ g_m overshoot.” However, this would be misleading; as explained, this effect has nothing to do with g_m , as defined in (8.2.4) for the *intrinsic* device. Thus, instead of having to rederive “effective” g_m values, it is best to just include R_{be} in the model for the complete device (Sec. 8.4); this will take care of the effects just discussed.

8.2.5 Weak Inversion

Using the weak inversion current equation (4.6.13) in the definition for g_m (8.2.4), we find

$$g_m = \frac{1}{n} \frac{I_{DS}}{\phi_t} \quad (8.2.41)$$

where n is given by (4.6.16), assuming the effect of interface traps is negligible, a valid assumption for many modern devices (Sec. 2.6). If this effect is not negligible, n can have a larger value, best determined experimentally.

Thus, in contrast to the behavior in strong inversion, here the quantity g_m/I_{DS} is independent of W/L . This is a consequence of I_{DS} being an almost exponential function of V_{GS} in weak inversion. In fact, the behavior here is qualitatively the same as for a bipolar transistor. For the latter, the current in the forward active region is given by $I_C \approx I_O e^{V_{BE}/\phi_t}$, where V_{BE} is the base-emitter voltage and I_O is a characteristic current independent of V_{BE} . Thus, $g_m = dI_C/dV_{BE} = I_C/\phi_t$, or $g_m/I_C = 1/\phi_t$, independent of geometrical details. Note that the corresponding quantity for the MOS transistor, $g_m/I_{DS} = 1/(n\phi_t)$, is always smaller since n is larger than unity. The value $1/\phi_t$, reached by the transconductance-to-current ratio for the bipolar transistor but not for the MOS transistor, is sometimes called the *Boltzmann limit*.

The substrate transconductance will not be obtained from (4.6.13) since I'_M , V_M , and n depend on V_{SB} in a complicated manner. We will use instead (4.6.6)–(4.6.7) with $V_{GB} = V_{GS} - V_{BS}$ and $V_{DB} = V_{DS} - V_{BS}$. Then from the definition of g_{mb} , (8.2.5), we obtain (Prob. 8.13)

$$g_{mb} \approx \frac{n-1}{n} \frac{I_{DS}}{\phi_t} \quad (8.2.42)$$

Thus the ratio g_{mb}/g_m is given by:

$$\frac{g_{mb}}{g_m} \approx n - 1 \quad (8.2.43a)$$

$$\approx \frac{\gamma}{2\sqrt{V_{SB} + 2\phi_F}} \quad (8.2.43b)$$

$$\approx \frac{\epsilon_s}{\epsilon_{ox}} \frac{t_{ox}}{d_B} \quad (8.2.43c)$$

where we have assumed negligible interface trap density, and where in the last expression d_B is the approximate width of the depletion region in weak inversion. This relation is of the same form as (8.2.20).

For g_{sd} we find, using the definition (8.2.6) in (4.6.13),

$$g_{sd} = \frac{e^{-V_{DS}/\phi_t}}{1 - e^{-V_{DS}/\phi_t}} \frac{I_{DS}}{\phi_t} \quad (8.2.44)$$

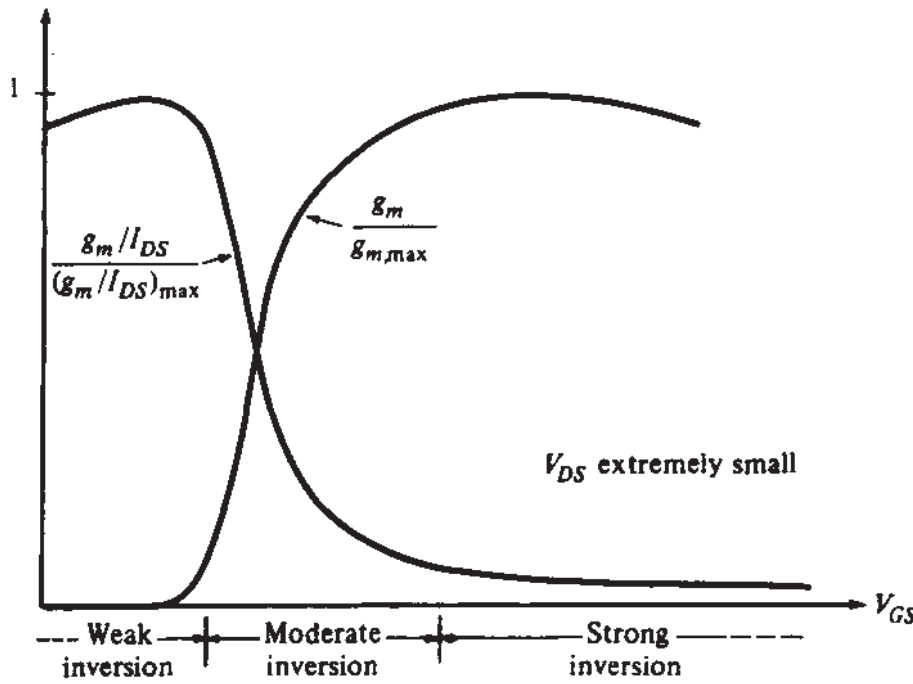
This equation predicts a g_{sd} that rapidly goes to 0 with increasing V_{DS} . However, this neglects the direct influence of the drain field on the channel (already discussed above for the case of strong inversion). Because of this effect, one often finds that for large V_{DS} a behavior similar to that predicted for saturation by our simple strong-inversion model is exhibited. One can again use an approximation of the form:

$$g_{sd} = \frac{I'_{DS}}{V_{AW}}, \quad V_{DS} > 5\phi_t \quad (8.2.45)$$

where V_{AW} is typically smaller than the corresponding quantity V_A used for strong inversion, and $V_{DS} > 5\phi_t$ has been imposed as a condition to ensure operation in the "flat" part of the curves in Fig. 4.25. For short-channel devices, g_{sd} can be larger than the value predicted by the above equation, as V_M in (4.6.13) stays close to the threshold voltage and can be drastically affected by DIBL. Working with (4.6.13) and assuming the changes of V_M with V_{DS} track those of V_T , it is easy to show that (8.2.34) is valid in weak inversion. This equation leads again to (8.2.35). From this and (8.2.41) it is seen that g_{sd} can still be written as in (8.2.45), with V_{AW} roughly proportional to L again (but with a smaller constant of proportionality).

8.2.6 Moderate Inversion

As has been noted in Sec. 4.5, in some models for I_{DS} , the moderate-inversion region is neglected, and the weak- and strong-inversion regions are taken to be adjacent. This results in some error in predicting I_{DS} , which can be tolerated for some applications. Unfortunately, though, the corresponding error in the small-signal parameters is usually unacceptable, and cannot be made small no matter what empirical values are chosen for the model parameters. To illustrate this, we will choose g_m as an example and will consider the behavior of this parameter over a wide range of V_{GS} for a fixed V_{DS} . To avoid the interference of higher-order effects, such as channel length modulation or uncertainty as to the exact value of V'_{DS} , we will not choose a V_{DS} in the saturation region. We will choose instead a *negligibly small* V_{DS} , which corresponds to a very common condition for measuring the threshold voltage and the quantity $\mu C'_{ox}$ (see Fig. 4.28). The results for g_m and for g_m/I_{DS} are shown in Fig. 8.11.⁴⁸ These results were obtained by using the general charge sheet model of Sec. 4.3.1, which is valid in all regions of operation (see also Sec. 8.2.7). As seen, in strong inversion g_m is approximately constant (g_m reaches a maximum and then decreases with increasing V_{GS} (due to the influence of V_{GS} on the effective mobility⁴⁹) corresponding to the slope of the solid line in Fig. 4.28.) In weak inversion, g_m/I_{DS} is seen to be approximately constant. This is expected from (8.2.41) and is a consequence of the practically exponential dependence of I_{DS} on V_{GS} . In moderate inversion, neither g_m nor g_m/I_{DS} can be considered approximately constant, as seen in the figure. Yet, models which neglect the moderate inversion region predict that the regions of constant g_m and of constant g_m/I_{DS} are adjacent at some point; clearly, *no* such point exists in Fig. 8.11. Thus, using such models in analog circuit applications and for V_{GS} values in the vicinity of V_T can lead to *very* wrong predictions.

**FIGURE 8.11**

Transconductance and transconductance-to-current ratio, both normalized to their maximum values, as a function of V_{GS} for negligibly small, fixed V_{DS} .⁴⁸

Unfortunately, there are as yet no analytically derived closed-form expressions for the small-signal conductances in the moderate-inversion region. Thus, one must resort to general models, discussed in the following subsection.

8.2.7 General Models

In Sec. 4.3 it was shown that general expressions can be developed for I_{DS} which are valid in all regions of operation for a long-channel device (neglecting channel length modulation). Such expressions can, in principle, be used to provide expressions for the small-signal conductances, which will be valid in all regions including moderate inversion and for the transition from nonsaturation to saturation in strong inversion.^{50–53} The difficulty here will, of course, be complexity. The drain current expression contains V_{GB} explicitly in (4.3.16) and (4.3.17), and also implicitly through ψ_{s0} and ψ_{sL} [see (4.3.18)]. As expected, the results are algebraically complicated. A different approach is now suggested. It can be shown[†] that for the charge sheet model, the quantities g_{sd} and g_{ss} , defined in (8.2.6) and (8.2.10), respectively, are given by⁵⁰

$$g_{sd} = \mu \frac{W}{L} (-Q'_{IL}) \quad (8.2.46)$$

$$g_{ss} = \mu \frac{W}{L} (-Q'_{I0}) \quad (8.2.47)$$

[†]The easiest way to derive this equation is through a quasi-Fermi potential formulation for the current; see (J.10) in Appendix J.

where Q'_{IL} and Q'_{IO} are the inversion layer charges per unit area at the drain end and at the source end of the channel, respectively. If desired, these can be expressed in terms of the surface potential by using (4.3.15). For example, g_{sd} becomes:

$$g_{sd} = \frac{W}{L} \mu C'_{ox} (V_{GB} - V_{FB} - \psi_{sL} - \gamma \sqrt{\psi_{sL}}) \quad (8.2.48)$$

As a check, consider the nonsaturation region. Substituting ψ_{sL} by $V_{DB} + \phi_0 = V_{DS} + V_{SB} + \phi_0$ in the above equation, produces (8.2.21). In the transition region from nonsaturation to saturation, the use of precise values for ψ_{sL} in (8.2.48) can provide accurate values for g_{sd} in long-channel devices. Of course, this equation should not be used in saturation, as it ignores channel length modulation and DIBL.

Simpler results are possible if the above relations are applied to the simplified charge sheet models discussed in Sec. 4.3.2. As an example, we will assume operation in saturation and will develop expressions valid in all regions of inversion. In saturation, Q'_{IL} is negligible. Using $Q'_{IL} = 0$ in (4.3.41b), solving this equation for Q'_{IO} in terms of I_{DS} , and using the result in (8.2.47) gives^{52,53} (Prob. 8.14):

$$g_{ss} = \frac{I_{DS}}{\phi_t} \frac{2}{1 + \sqrt{4 \frac{I_{DS}}{I_Z} + 1}}, \quad \text{saturation} \quad (8.2.49)$$

where

$$I_Z = \frac{W}{L} \mu C'_{ox} (2n\phi_t^2) \quad (8.2.50)$$

is a characteristic current.⁵⁴⁻⁵⁷ Further manipulation of the equations of the simplified charge sheet model produces^{52,53}

$$g_m \approx \frac{g_{ss}}{n}, \quad \text{saturation} \quad (8.2.51)$$

Since, in the saturation region, g_{sd} is negligible compared to g_m and g_{mb} (we assume no short-channel effects), (8.2.11) gives

$$g_m + g_{mb} \approx g_{ss}, \quad \text{saturation} \quad (8.2.52)$$

Using (8.2.51) and (8.2.52) we obtain

$$g_{mb} \approx \frac{n-1}{n} g_{ss}, \quad \text{saturation} \quad (8.2.53)$$

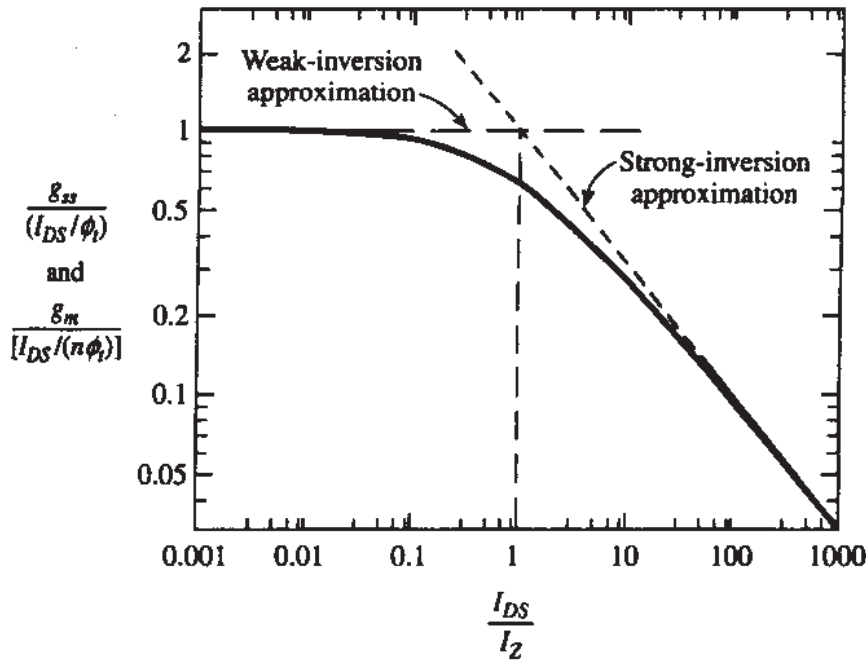


FIGURE 8.12

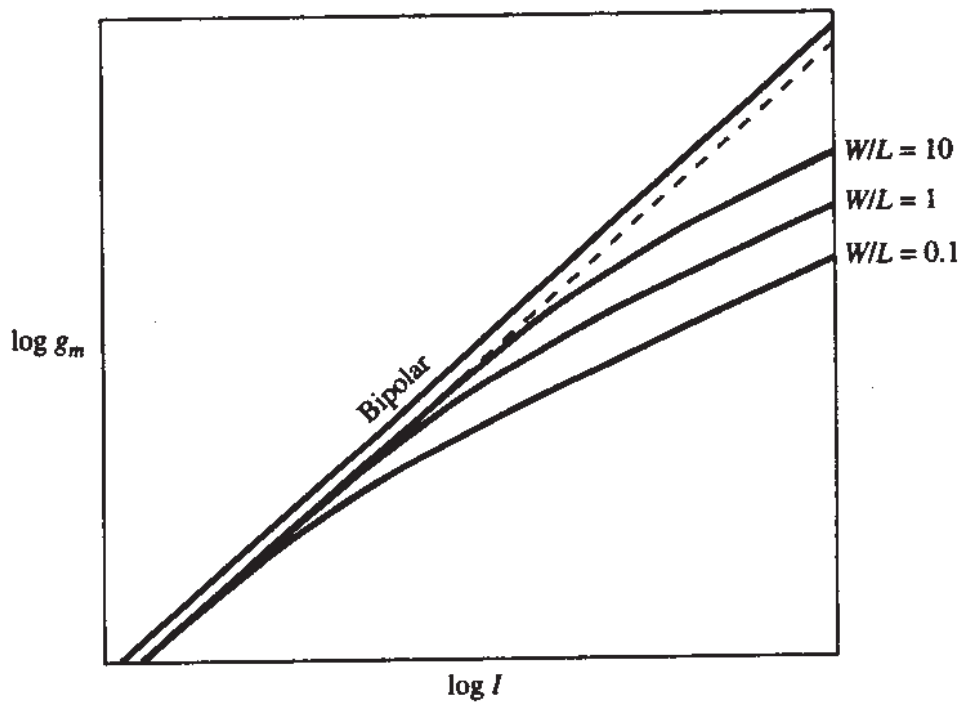
Source conductance and gate transconductance, normalized to their maximum values, versus normalized drain-source current.^{53,55,57,57a}

As a check, consider $I_{DS} \ll I_Z$; then the above expressions for g_m and g_{mb} reduce to those obtained for weak inversion in (8.2.41) and (8.2.42). Or consider $I_{DS} \gg I_Z$; then g_m can be easily seen to reduce to the strong inversion expression (8.2.17b), with n replacing α (the two quantities are of similar magnitude); similarly, g_{mb} approaches the value given by (8.2.20a). Similar results have been produced earlier using interpolation techniques.^{54–57}

The above results are plotted, in normalized form, in Fig. 8.12.^{53,55,57,57a} In the same plot, we show g_m plots from our earlier weak- and strong-inversion expressions. The point $I_{DS}/I_Z = 1$ falls in the moderate-inversion region. Clearly, the weak- and strong-inversion expressions fail for a range of about 2 orders of magnitude centered around this point. Instead, (8.2.49) has been extensively compared to measurements and has been found to be very accurate⁵³ throughout, for a variety of processes. This means⁵³ that, for a given temperature, g_{ss}/I_{DS} is fully determined by the ratio I_{DS}/I_Z . Note that the maximum possible value of g_{ss} is I_{DS}/ϕ_t (the “Boltzmann limit”), and that of g_m is $I_{DS}/(n\phi_t)$.

Thus, in saturation, for a given bias current in any region of inversion, the determination of the transconductances becomes trivial! One need only to find g_{ss} from (8.2.49) (or Fig. 8.12), and then find what part of g_{ss} is apportioned to each of g_m and g_{mb} , by using (8.2.51) and (8.2.53). In a similar manner, other small-signal parameters can also be expressed in terms of I_{DS}/I_Z .^{55,57} An example is shown in Table 8.1 at the end of this chapter.

In Fig. 8.13, $\log g_m$ is plotted versus $\log I_{DS}$ for various values of W/L and for operation in the flat part of the I_{DS} - V_{DS} characteristics; a fixed V_{SB} is assumed. For each W/L value, the straight-line segment on the *right* corresponds to strong inversion,

**FIGURE 8.13**

Comparison of $\log g_m$ versus $\log I$ behavior for a bipolar transistor and for MOS transistors with various W/L values, with fixed V_{SB} .

and the straight-line segment on the *left* corresponds to weak inversion. The curved part corresponds to moderate inversion. In strong inversion, for a given value of I_{DS} , one can obtain a larger g_m by using a larger value for W/L , as expected from (8.2.17b). However, this is not true in weak inversion, as can be seen both in the figure and from (8.2.41). There, once I_{DS} is fixed, g_m is known (assuming a fixed value for n). This behavior is compared to that of a bipolar device in Fig. 8.13 (see comments on the latter in Sec. 8.2.5).

Plots of g_m , g_{mb} , and g_{sd} vs. V_{GS} for a fixed, large V_{DS} have the form shown in Fig. 8.14. Such plots show these quantities varying continuously over all regions of inversion. The order in which the various regions are encountered as V_{GS} is increased has been discussed in Sec. 7.4.8 in conjunction with Fig. 7.8.

8.3 A MEDIUM-FREQUENCY SMALL-SIGNAL MODEL FOR THE INTRINSIC PART

8.3.1 Introduction

When the small-signal voltages applied to an MOS transistor are varying fast, the small-signal terminal currents can be very different from those found using the circuit of Fig. 8.3. We now introduce a simple model that will predict such behavior as long as the frequency is not too high. The model achieves this by taking into account charge storage effects, assuming quasi-static operation.^{1,2,5,8,13,15,16,28,29,58-97,51-57,42-44} Not all such effects are included; additional charge storage effects will be considered in

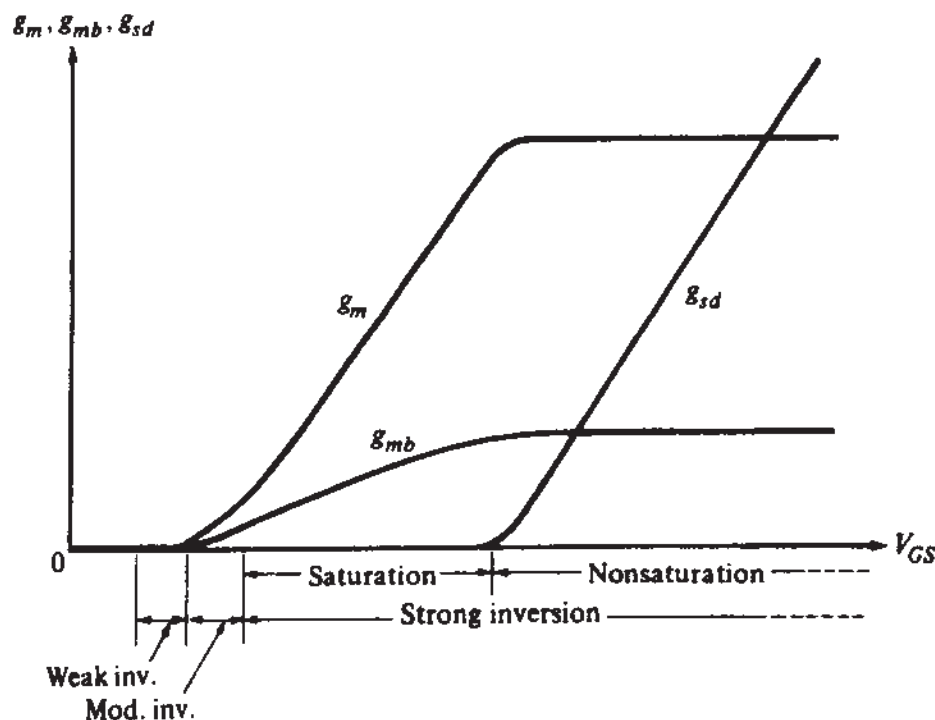


FIGURE 8.14

Small-signal gate transconductance, substrate transconductance, and drain conductance as a function of V_{GS} for fixed V_{DS} and V_{SB} as obtained from the complete charge sheet model, for a long-channel device.

Chap. 9. Nevertheless, the model we are about to introduce here is important in its own right, and we present it in a self-contained manner for two reasons:

1. It is widely in use, because it achieves a reasonable balance between accuracy and complexity. Some readers might find it adequate for their purposes, so it was felt that they should not have to be subjected to the general treatment of Chap. 9 before they encountered a useful model.
2. The material in this section will be a useful foundation for the general treatment in Chap. 9. In fact, the more complete model in that chapter can be produced just by adding a few extra elements to the model we will develop here.

As before, we will consider only the *intrinsic* part of the transistor in this section. This fact will be understood without being indicated explicitly. Extrinsic device modeling will be introduced in Sec. 8.4. No impact ionization is assumed.

In this section we will emphasize intuition. For more rigor, the reader will have to wait until Chap. 9.

8.3.2 Intrinsic Capacitances

The intrinsic part of a long-channel transistor (neglecting the two-dimensional effects very close to the source and drain regions) can be considered by itself by viewing it as a fictitious device in which the length of the source and drain regions has shrunk to zero, as shown in Fig. 8.15a. Our discussion will be more convenient if we define the

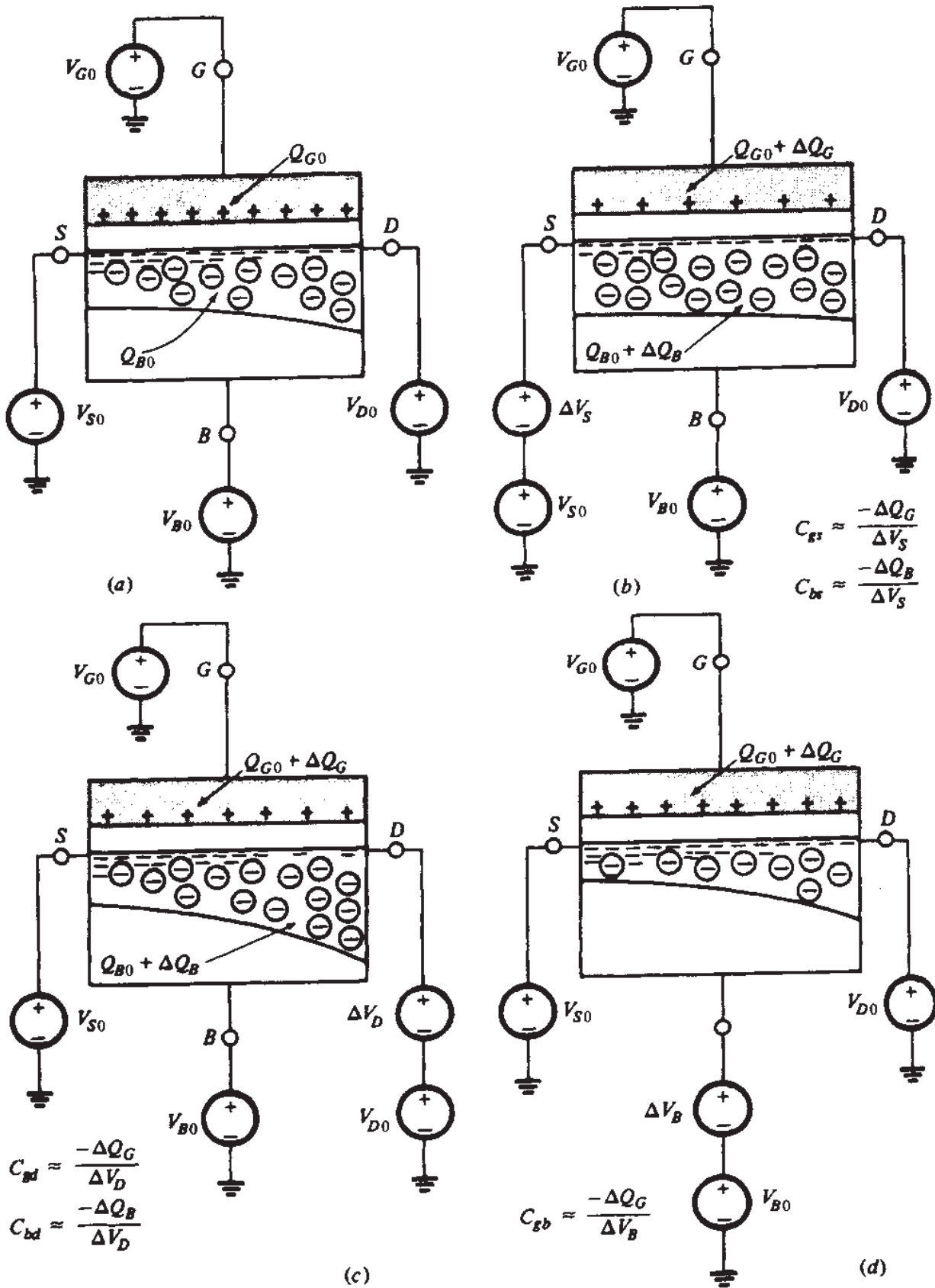


FIGURE 8.15

Measurement of intrinsic capacitances (in principle). The figure shows the intrinsic part of a transistor. (a) Transistor biased with four dc voltages at a certain operating point; (b) effect of a small increase in the source voltage; (c) effect of a small increase in the drain voltage; (d) effect of a small increase in the substrate voltage. The relations in the figures give the corresponding capacitance values.

voltages at the four terminals with respect to an arbitrary reference, as shown by the “ground” symbol in the figure. A subscript 0 will be used to denote the “quiescent” values of the voltages and charges in Fig. 8.15a. Let us consider the effects of small changes of V_S on the gate and depletion region charges (Fig. 8.15b), the effect of small changes of V_D on the gate and depletion region charges (Fig. 8.15c), and the effect of small changes of V_B on the gate charge (Fig. 8.15d). The details in these figures and the definitions next to them will be explained shortly. In order to discuss a concrete case, and to be able to draw the figures, we have assumed that the device operates in strong inversion, and that ΔV_S , ΔV_D , and ΔV_B represent increases. However, the definitions given are general; they apply in all regions, and are independent of the sign of the voltage increments. All voltages are assumed constant before and after the change, and all cases in Fig. 8.15 are assumed to be in dc steady state. Thus, for example, ΔQ_G in Fig. 8.15b is the difference between the total steady-state charge when the source voltage is held fixed at $V_S + \Delta V_S$ and the steady-state charge when the source voltage is held fixed at V_S . For simplicity, the interface charge Q_o is not shown.

We now consider the experiment of Fig. 8.15b in detail. The potential across the oxide at various points decreases relative to that in Fig. 8.15a and, hence, the gate charge decreases too, and thus ΔQ_G is negative. The relation between the cause (ΔV_S) and the effect (ΔQ_G) can be represented by the small-signal equivalent circuit of Fig. 8.16. In this circuit the voltage ΔV_S places a charge $C_{gs} \Delta V_S$ on the bottom plate and a charge $-C_{gs} \Delta V_S$ on the top. For the latter charge to represent the change ΔQ_G in Fig. 8.15b, we must have $-C_{gs} \Delta V_S = \Delta Q_G$. Hence $C_{gs} = -\Delta Q_G / \Delta V_S$, as stated next to Fig. 8.15b. More formally, we have

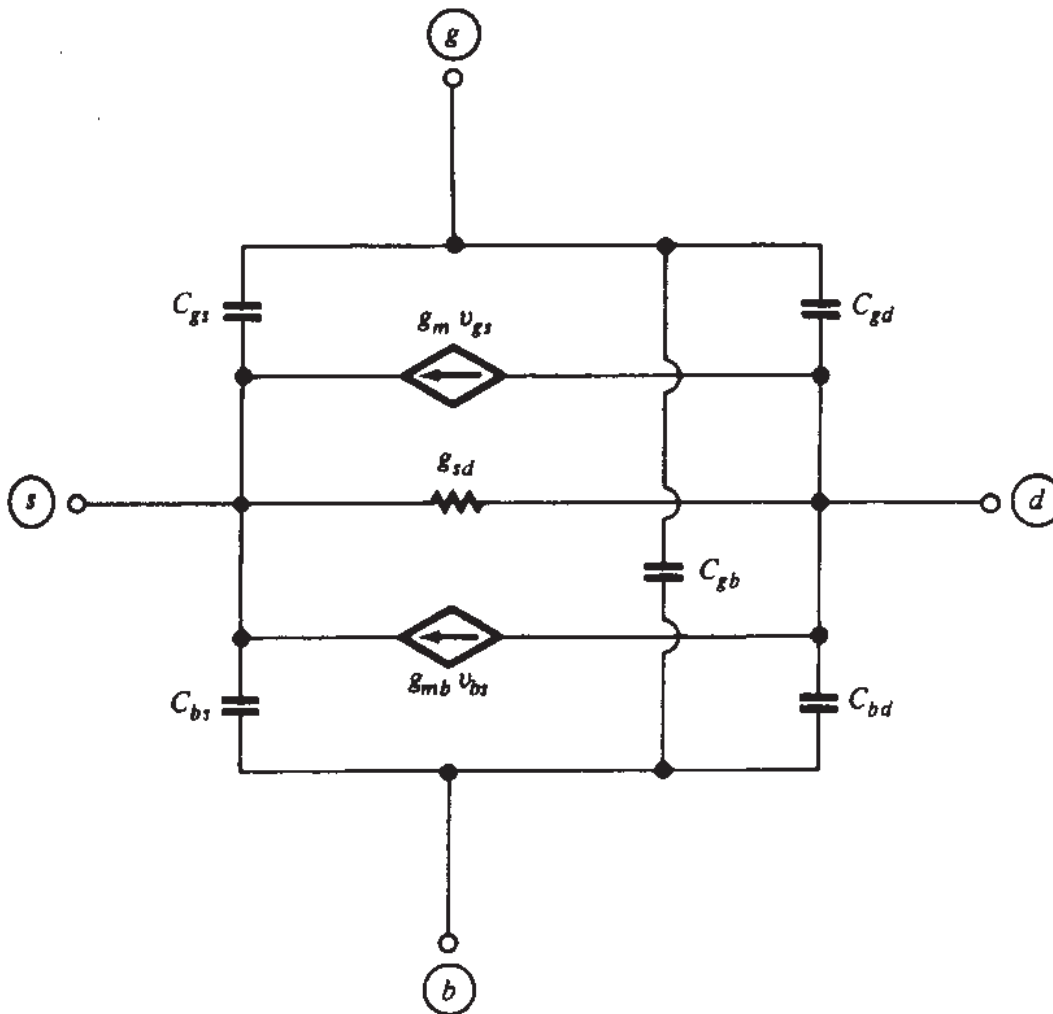
$$C_{gs} = - \left. \frac{\partial Q_G}{\partial V_S} \right|_{V_G, V_D, V_B} \quad (8.3.1)$$

where to the right of the vertical line we show the voltages being held constant. The reader is urged to carefully consider the rationale given for the minus sign in the above definition.

It is important *not* to associate C_{gs} with any parallel plate structure in Fig. 8.15. C_{gs} is simply the value that the fictitious capacitor in Fig. 8.16 should have in order for the charge on its top plate to be the same as the charge *change* ΔQ_G in Fig. 8.15b. Note that, since ΔQ_G is negative when ΔV_S is positive, the value of C_{gs} is positive.†

Figure 8.15b also illustrates the capacitive effect of the source on the substrate. Increasing V_S increases the width of the depletion region, thus making the total charge there more negative, and hence ΔQ_B is negative. This is accomplished by more acceptor atoms being uncovered (in comparison to Fig. 8.15a), which means that holes must leave the bottom of the depletion region, which causes a total charge

†Throughout this section, all capacitances and charges are for the intrinsic part of the device only. For example, C_{gs} is the *intrinsic* gate-source capacitance and *not* the total gate-source capacitance of the device. It will be seen in Sec. 8.4 that an extrinsic capacitance caused by the physical proximity of the gate and the source must be added to C_{gs} to arrive at the total gate-source capacitance.

**FIGURE 8.17**

Simple small-signal equivalent circuit for the MOS transistor.

All the above five effects can be modeled in a manner analogous to that of Fig. 8.16. These effects can be included in a small-signal equivalent circuit by adding to the circuit of Fig. 8.3 five capacitors, as shown in Fig. 8.17. It is important to note that the various elements in the resulting model *do not interfere with each other*. For example, keeping terminals g , d , and b at ground and applying ΔV_s at terminal s will cause a charge $-C_{gs} \Delta V_s$ to enter terminal g , thus modeling correctly the effect of the source on the gate. C_{gd} and C_{gb} , although connected to the gate, will *not* interfere since they will act as open circuits (the voltage across them is fixed). The reader can easily check that each of the remaining four capacitance effects is also modeled correctly by the corresponding capacitors and with no interference from other elements (Prob. 8.15). This “noninterference” is a very important property which should be carefully checked every time several elements—each meant to simulate a separate effect—are connected together to form a circuit model.

Since C_{gs} , C_{gd} , and C_{gb} represent the effect on the gate of each of the other three terminals, the circuit of Fig. 8.17 can model accurately the quasi-static small-signal effects one “sees looking into” the gate of a MOS transistor. Notice that for now we cannot prove a similar statement for the other terminals. (This subject will be discussed further in Chap. 9.) At this point, we will only say that, as it turns out by comparison to more

complete models (Chap. 9), the model topology of Fig. 8.17 is satisfactory (as far as looking into any terminal, and as far as the effect of any terminal on any other is concerned), even if the terminal voltages vary continuously with time, as long as their variation is slow enough. For sinusoidal small variations, this can be quantified by establishing an upper frequency limit of validity. The figure one comes up with depends on the accuracy desired, the region of operation, which terminals are being driven with signals, which terminal currents are of interest, whether the magnitude or the phase of those currents is of interest,[†] etc. Thus, it is not easy to give a single number. Some general indication can, nevertheless, be given by comparison to more sophisticated model topologies (Chap. 9). Thus, in strong inversion, whatever the criterion being used, the upper frequency limit of validity turns out to be proportional (but *not* equal) to the quantity:

$$\omega_o = \frac{\mu(V_{GS} - V_T)}{\alpha L^2} \quad (8.3.6)$$

where the reason for the presence of the quantities in the right-hand side can be understood after higher-order models are discussed in Chap. 9. Thus, assume that satisfactory values have been obtained for all model elements at low frequencies. Then, without touching these element values, the model will continue to be valid up to about $0.1\omega_o$ (a conservative limit, for very critical applications) or even $0.5\omega_o$ (for noncritical applications). Such performance is sufficient in many cases. The degradation of the model is very gradual as the frequency is increased, so no “sharp” deterioration is observed at any particular frequency. When, at high frequencies, the model eventually becomes unacceptable, it will not be because the element values are not right, but rather because the *nature* of the model (Fig. 8.17) is inadequate for such frequencies. The only way to achieve satisfactory modeling in this case is to use more sophisticated model topologies, as discussed in Chap. 9.

The five capacitances defined above are strongly dependent on the “bias values” of the terminal voltages, i.e., the values around which the small-signal voltages occur. These bias values are denoted by V_{D0} , V_{G0} , V_{B0} , and V_{S0} in Fig. 8.15a. For simplicity, from now on we will denote them simply as V_D , V_G , V_B , and V_S . We now present expressions for the capacitances in terms of the bias voltages. We will consider each region of operation separately.

STRONG INVERSION. Accurate capacitance expressions can be derived for the strong-inversion region by using the above capacitance definitions in conjunction with the charges corresponding to the complete model of Sec. 4.5.1. Expressions for these charges are given in Appendix M. However, the resulting capacitance expressions are complicated.^{13,15‡}

[†]As the frequency is raised, both the magnitude and the phase as predicted by the model will be increasingly in error. For many applications, the phase error becomes intolerable first.

[‡]For this reason, empirical capacitance expressions are often used in conjunction with the complete strong-inversion model (Appendix M).

For simpler, approximate results, one can use the charge expressions corresponding to the simplified model of Sec. 4.5.3. Depending on the generality one allows for the value and functional dependence of α (Sec. 4.5.3), the resulting capacitance expressions can range from simple to very complicated. A good compromise between simplicity and accuracy can be obtained based on the following two assumptions: (1) the value of α is equal to

$$\alpha_1 = 1 + \frac{\gamma}{2\sqrt{\phi_0 + V_{SB}}} = 1 + \frac{dV_T}{dV_{SB}} \quad (8.3.7)$$

which was one of the choices for α in Sec. 4.5.3, and (2) the derivative of α_1 with V_S and V_B is negligible; i.e., in our differentiation of the charge expressions, the quantity α_1 will be treated as a *constant*. These assumptions simplify the calculations, because every time a charge containing the term V_T is differentiated with respect to V_{SB} , it will result in the term $1 + dV_T/dV_{SB} = \alpha_1$. This term can then be grouped together with other terms containing α_1 , resulting in simple expressions. From the material in Sec. 4.5.3, it is clear that the above assumptions can be justified especially for large V_{SB} and small V_{DS} (or small V_{GS} , since then V'_{DS} will be small). Hence, the accuracy of the resulting capacitance expressions can be expected to be good in such cases. In cases where V_{DS} is large and/or V_{SB} is small, some error can be expected. However, such error is consistent with the overall accuracy of the simplified model. We will later show how the above error can be reduced.

Using then $\alpha = \alpha_1$ in (7.4.15) and (7.4.17), applying the definitions (8.3.1) to (8.3.5), and neglecting the derivative of α_1 with V_S and V_B , we obtain, after much algebra (Prob. 8.16):

$$C_{gs} = C_{ox} \frac{2(1 + 2\eta)}{3(1 + \eta)^2} \quad (8.3.8)$$

$$C_{bs} = (\alpha_1 - 1)C_{ox} \frac{2(1 + 2\eta)}{3(1 + \eta)^2} \quad (8.3.9a)$$

$$= (\alpha_1 - 1)C_{gs} \quad (8.3.9b)$$

$$C_{gd} = C_{ox} \frac{2(\eta^2 + 2\eta)}{3(1 + \eta)^2} \quad (8.3.10)$$

$$C_{bd} = (\alpha_1 - 1)C_{ox} \frac{2(\eta^2 + 2\eta)}{3(1 + \eta)^2} \quad (8.3.11a)$$

$$= (\alpha_1 - 1)C_{gd} \quad (8.3.11b)$$

$$C_{gb} = \frac{\alpha_1 - 1}{3\alpha_1} C_{ox} \left(\frac{1 - \eta}{1 + \eta} \right)^2 \quad (8.3.12)$$

where C_{ox} is the total *intrinsic* oxide capacitance:

$$C_{ox} = C'_{ox}WL \quad (8.3.13a)$$

and η is the parameter defined in (4.5.38) and plotted in Fig. 4.20.

In the case of C_{bs} and C_{bd} we note that, although it is convenient to express them in terms of $(\alpha_1 - 1)C_{ox}$, this quantity is *independent* of C_{ox} . This is because, from (8.3.7), $(\alpha_1 - 1)$ is proportional to γ which, from (2.5.16), is inversely proportional to C'_{ox} . Thus we have:

$$(\alpha_1 - 1)C_{ox} = \frac{\gamma}{2\sqrt{\phi_0 + V_{SB}}} C'_{ox}WL = \frac{\sqrt{2q\epsilon_s N_A}}{2\sqrt{\phi_0 + V_{SB}}} WL = C'_{bc}(V_{SB})WL \quad (8.3.13b)$$

where $C'_{bc}(V_{SB})$ is the substrate-channel depletion region capacitance per unit area, at the source end of the channel [similar to (1.5.19)].

The five capacitances are plotted in Fig. 8.18 using the above expressions. The behavior seen will be discussed intuitively shortly.

Comparisons to (complicated) results⁶² from the charge sheet model (or from the complete strong inversion model—see Appendix M) show that the accuracy of the expressions developed above for C_{gs} and C_{gd} is excellent for all V_{DS} . The expressions for C_{bs} , C_{bd} , and C_{gb} are found to be accurate at $V_{DS} = 0$. When V_{GS} and V_{DS} are large, though, and at the same time V_{SB} is small, the accuracy is not good. Nevertheless, in many circuit applications the consequences of the above errors will not be serious because C_{bs} , C_{bd} , and C_{gb} are usually small and are in parallel with other capacitances (e.g., extrinsic capacitances and the capacitances of other devices in the circuit). C_{bs} in particular is often short-circuited out through a source-substrate connection. One should also point out that within the above errors the expressions for the three capacitances track satisfactorily large variations in oxide thickness and substrate doping concentration, which is a desirable property for any model parameter. If more accuracy is desired, one can use for α_1 a simple modification. Thus, if in the expression for C_{gb} we substitute α_1 by

$$\alpha_s = 1 + \frac{\gamma}{2\sqrt{\phi_0 + V_{SB} + k_c V'_{DS}(1 - \alpha)}} \quad (8.3.14)$$

with $k_c = 1$, the accuracy is much improved. Similarly, the accuracy of C_{bs} and C_{bd} can be improved if we substitute α_1 by the above, with k_c a small number (e.g., 0.1 to 0.2).

From (8.3.8) to (8.3.11) and (8.2.14) we obtain the following interesting approximate result, mostly valid at small V_{DS} or low V_{GS} values:⁵⁸

$$\frac{C_{bs}}{C_{gs}} \approx \frac{C_{bd}}{C_{gd}} \approx \frac{C'_{bc}(V_{SB})}{C'_{ox}} \approx \frac{g_{mb}}{g_m} \approx \frac{dV_T}{dV_{SB}} = \alpha_1 - 1 \quad (8.3.15)$$

This result becomes exact for $V_{DS} = 0$.

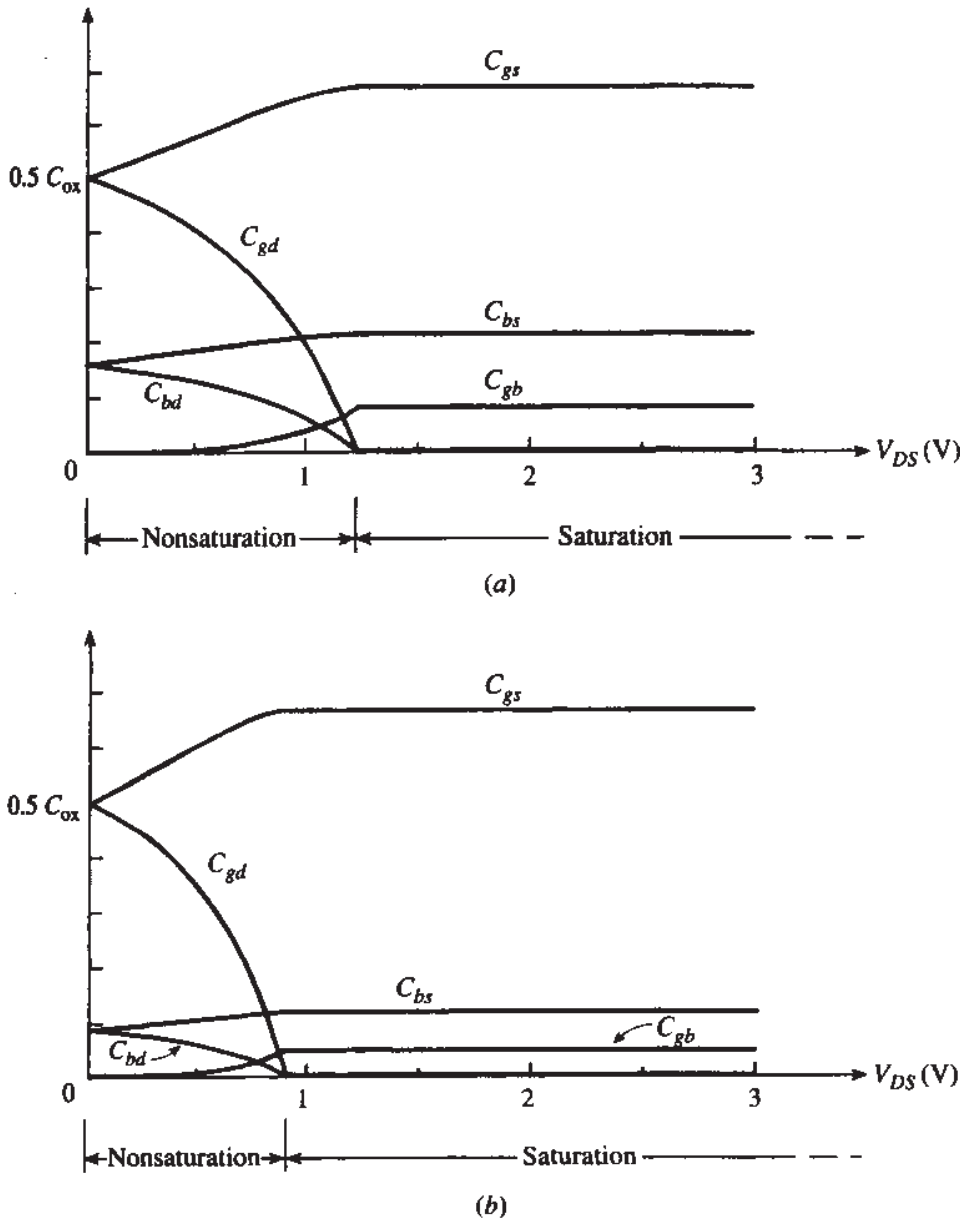


FIGURE 8.18 Small-signal capacitances versus V_{DS} for a device with $V_{T0} = 0.5$ V, $\gamma = 0.6$ V^{1/2}, $\phi_0 = 0.9$ V, with $V_{GS} = 2$ V, as obtained from (8.3.8) to (8.3.12). (a) $V_{SB} = 0$; (b) $V_{SB} = 2$ V.

We now consider the capacitances for two cases of interest.

NONSATURATION WITH $V_{DS} = 0$. With $V_{DS} = 0$ ($\eta = 1$), the capacitance expressions give:

$$C_{gs} = C_{gd} = \frac{C_{ox}}{2} \quad (8.3.16)$$

$$C_{bs} = C_{bd} = (\alpha_1 - 1) \frac{C_{ox}}{2} = \frac{1}{2} C'_{bc}(V_{SB})WL \quad (8.3.17)$$

$$C_{gb} = 0 \quad (8.3.18)$$

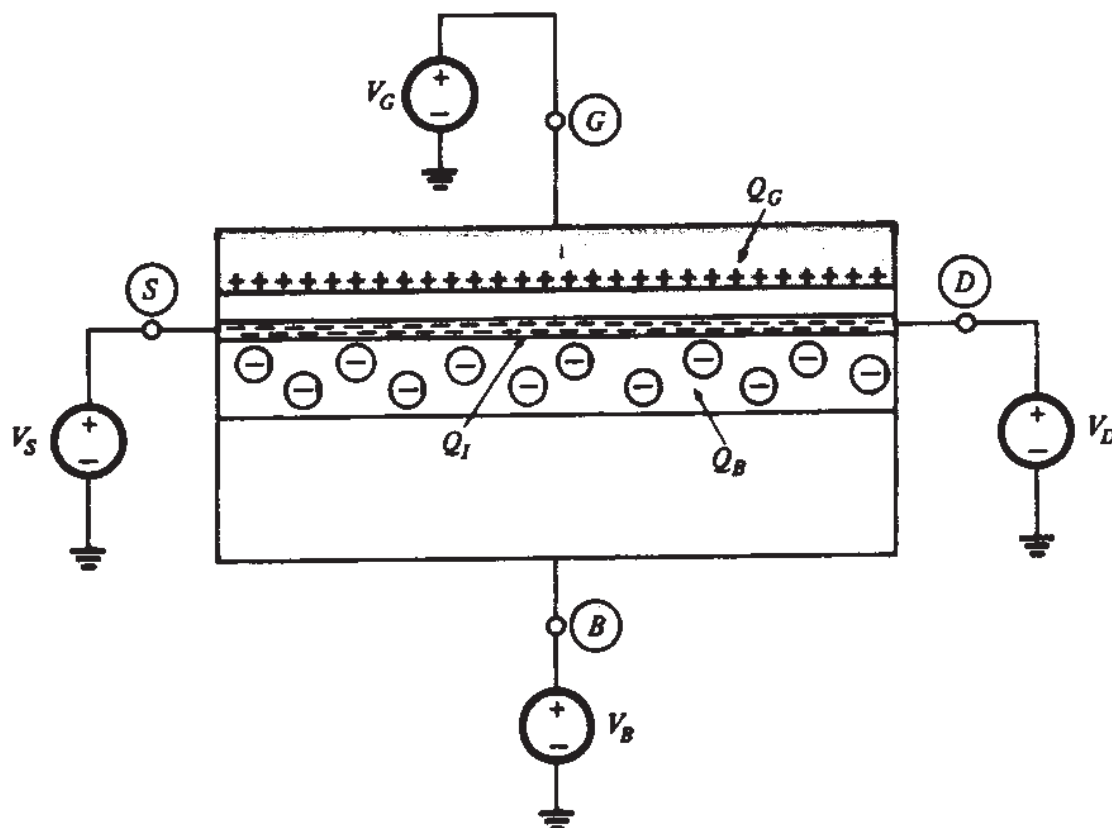


FIGURE 8.19

A MOS transistor with $V_D = V_S$.

It is easy to make these results intuitively plausible. With $V_{DS} = 0$ the channel is as shown in Fig. 8.19. Practically all the gate field lines terminate on the inversion layer, which is assumed strongly inverted throughout. If V_G and V_B are kept fixed, and both V_S and V_D are increased by the same amount ΔV , the potential across the oxide will decrease at every point by ΔV . Hence, the corresponding gate charge decrease will be $C_{ox} \Delta V$. If, instead, only V_S is increased by ΔV whereas V_D is kept fixed, the potential *change* across the oxide will vary from ΔV at the source end to 0 at the drain end. This variation will be linear for a very small ΔV . It is easy to see that now the decrease in the gate charge will only be half as much as before, i.e., $\frac{1}{2} C_{ox} \Delta V$ (Prob. 8.17). Thus, the value of C_{gs} needed to properly model this effect is $\frac{1}{2} C_{ox}$. A similar argument can be given for C_{gd} , thus verifying (8.3.16).

The values of C_{bs} and C_{bd} can be discussed in a similar manner. Here the arguments used above apply with C_{ox} replaced by $C'_{bc}(V_{SB})WL$ which, for the case we are considering, is seen to be the total capacitance of the uniform reverse-biased "field-induced junction" formed by the inversion layer and the substrate in Fig. 8.19. Thus (8.3.17) also makes sense.

We will now attempt to make the relation to transconductances, shown in (8.3.15), intuitively plausible. Consider a nonzero, very small V_{DS} , so that we can have a small nonzero current I_{DS} (g_m and g_{mb} would vanish if V_{DS} were 0). As long as V_{DS} is very small, it will not upset appreciably the uniformity in the channel, and the

picture will be practically as shown in Fig. 8.19. Keeping now V_B , V_S , and V_D fixed, let V_G change by ΔV_G . This will cause an inversion layer charge change of magnitude

$$|\Delta Q_{I1}| = (C'_{ox} WL) \Delta V_G \quad (8.3.19)$$

and a corresponding current change

$$\Delta I_{D1} = g_m \Delta V_G \quad (8.3.20)$$

Repeating this experiment by using the “back gate” instead of the front gate, we obtain

$$|\Delta Q_{I2}| = [C'_{bc}(V_{SB}) WL] \Delta V_B \quad (8.3.21)$$

$$\Delta I_{D2} = g_{mb} \Delta V_B \quad (8.3.22)$$

Since the inversion layer is uniform, the current is proportional to Q_I . Thus $\Delta I_{D2}/\Delta I_{D1} = |\Delta Q_{I2}|/|\Delta Q_{I1}|$. Substituting the quantities in this relation by the corresponding right-hand sides in (8.3.19) to (8.3.22), we obtain $g_{mb}/g_m = C'_{bc}(V_{SB})/C'_{ox}$. This, with the use of (8.3.13b), (8.3.9), and (8.3.11) agrees with (8.3.15).

We now consider C_{gb} . At $V_{DS} = 0$, the strongly inverted electron layer, which is “connected” to the fixed voltages $V_S = V_D$, keeps the voltage across the oxide fixed at all points, even if V_B is varied. Hence the gate does not feel the variation, and $\Delta Q_G = 0$. This is modeled by [see (8.3.5)]

$$C_{gb} = 0, \quad V_{DS} = 0 \quad (8.3.23)$$

In other words, the strong-inversion layer acts as a “shield” throughout the channel, protecting the gate from the influence of the substrate.

It is interesting to point out that, for $V_{DS} = 0$, complicated capacitance expressions corresponding to the complete model reduce precisely to (8.3.16) to (8.3.18). It is easy to trace this full agreement to our choice of value for α . Indeed, $\alpha = \alpha_1$ was seen in Sec. 4.5.3 to be the best choice for $V_{DS} = 0$.

SATURATION. In the saturation region ($\eta = 0$), we have from (8.3.8) to (8.3.12):

$$C_{gs} = \frac{2}{3} C_{ox} \quad (8.3.24)$$

$$C_{bs} = \frac{2}{3} (\alpha_1 - 1) C_{ox} \quad (8.3.25)$$

$$C_{gd} = 0 \quad (8.3.26)$$

$$C_{bd} = 0 \quad (8.3.27)$$

$$C_{gb} = \frac{\alpha_1 - 1}{3\alpha_1} C_{ox} \quad (8.3.28)$$

The results for C_{gd} and C_{bd} can be easily seen to make sense as follows. In the saturation region, communication from the drain to the rest of the device is cut off owing to pinchoff (neglecting channel length modulation). Hence, when V_D is varied, the intrinsic device is not affected and all charges in it remain the same. In particular, $\Delta Q_G = 0$ and $\Delta Q_B = 0$. This, from (8.3.3) and (8.3.4) implies that $C_{gd} = 0$ and $C_{bd} = 0$, in agreement with (8.3.26) and (8.3.27).

Consider now C_{gs} in (8.3.24). This value is accurate, as found by comparisons with accurate models and measurements (see below). The fact that C_{gs} is found to be less than C_{ox} often comes as a surprise, because one tends to think of the inversion layer as the lower of two "parallel plates" of a capacitor, the other plate being the gate. If we really did have a parallel metal plate all under the oxide (neglecting the pinchoff region near the drain), and if it were attached to the source and cut off from the drain, then we would, in fact, obtain a capacitance equal to C_{ox} . This is because, if the source potential were changed by ΔV_S , we would have $|\Delta\psi_{ox}(x)| = |\Delta V_S|$ for any position x along the channel, with ψ_{ox} the oxide potential. However, this is *not* what happens in the transistor. Using the analysis in Sec. 4.5.3 and (3.2.2), one can show that we will have $|\Delta\psi_{ox}| = |\Delta V_S|$ *only* at the source end of the channel, and that $|\Delta\psi_{ox}(x)|$ will be smaller and smaller as one goes toward the drain (in fact, for the complete model of Sec. 4.5.1, one can show that, in saturation, ψ_{ox} at the *drain* end remains fixed at a value corresponding to the gate-drain threshold voltage, which is independent of V_S). Thus, points further away from the source feel the change in the source potential less and contribute less to the change in the gate charge. This is why C_{gs} is less than C_{ox} .†

Let us now consider C_{bs} in saturation. Notice that (8.3.25) is of the form of (8.3.24). One can obtain an intuitive feeling for this result for the case of large V_{SB} and small V'_{DS} . In such cases, the effective reverse bias of the channel with respect to the substrate does not vary much along the channel. Thus, the depletion region width is roughly uniform, just as was the case with $V_{DS} = 0$, and one can argue as we did above for C_{gs} , with the role of C'_{ox} played now by $C'_{bc}(V_{SB})$. Thus, the result in (8.3.25) is not surprising. If the above conditions on bias voltages are not satisfied, though, the above intuition does not hold and, in fact, the model is not very accurate, as has been seen. If V_{GS} and, thus, V'_{DS} are large, the depletion region width will increase significantly as we go toward the drain, making the influence of the channel on the substrate weaker (as V_S is varied). C_{bs} is then found to be somewhat *less* than the value given by (8.3.25), as has been seen.

Finally, we consider C_{gb} . Assume a change ΔV_B of V_B in Fig. 8.15d. Working as explained in the preceding paragraph, we find that here the resulting change in the

†In circuits literature, the value $\frac{2}{3} C_{ox}$ in (8.3.24) is sometimes justified by saying that the length of the pinch-off region is $\frac{1}{3} L$, and thus the effective channel length is $\frac{2}{3} L$. This explanation is *not* correct. The factor $\frac{2}{3}$ is the result of a mathematical derivation which has nothing to do with the length of the pinchoff region. This length, in fact, must be assumed zero in order to arrive at (8.3.8) and, thus, at (8.3.24). Besides, the length of the pinchoff region is not $\frac{1}{3} L$ in general; in fact it is predicted from first-order theory to be independent of L from (6.2.1). For long-channel devices, the pinchoff region occupies a very small part of the channel.

oxide potential, $\Delta\psi_{ox}$, although zero at the source end, becomes nonzero at other points, and in fact $|\Delta\psi_{ox}(x)|$ increases toward the drain. (For example, for the complete model of Sec. 4.5.1 in saturation, ψ_{ox} at the drain is maintained at a value corresponding to the gate-drain threshold, as already mentioned; when V_B is changed, this value changes due to the body effect at that point.) The nonzero changes in $\psi_{ox}(x)$ contribute to a change in the gate charge. Thus C_{gb} is nonzero in saturation, as predicted by (8.3.28).

Closing our discussion of strong-inversion capacitances, we should point out that the expressions developed above, although derived by using the approximate model charges, are sometimes used in conjunction with other models. Obviously, in such cases, one should use in the expression for η a value of V'_{DS} as predicted by such models, to be consistent with the I_{DS} - V_{DS} characteristics predicted by them.

INTRINSIC TRANSITION FREQUENCY. Consider a transistor in the connection shown in Fig. 8.20a, where the bias is assumed such that operation is in the saturation region. The voltage $\epsilon \sin \omega t$ is a sinusoidal small signal of angular frequency ω (in rad/s). The small-signal equivalent circuit for this connection can be derived by substituting the transistor by the circuit of Fig. 8.17, and by substituting all dc voltage sources with short circuits (since for these sources $\Delta V = 0$). Removing now all elements which appear in parallel with short circuits and noting that C_{gd} is 0 in saturation, we arrive at the circuit of Fig. 8.20b. The small-signal drain and gate currents can be calculated by using this circuit, and are shown directly on the figure. Defining

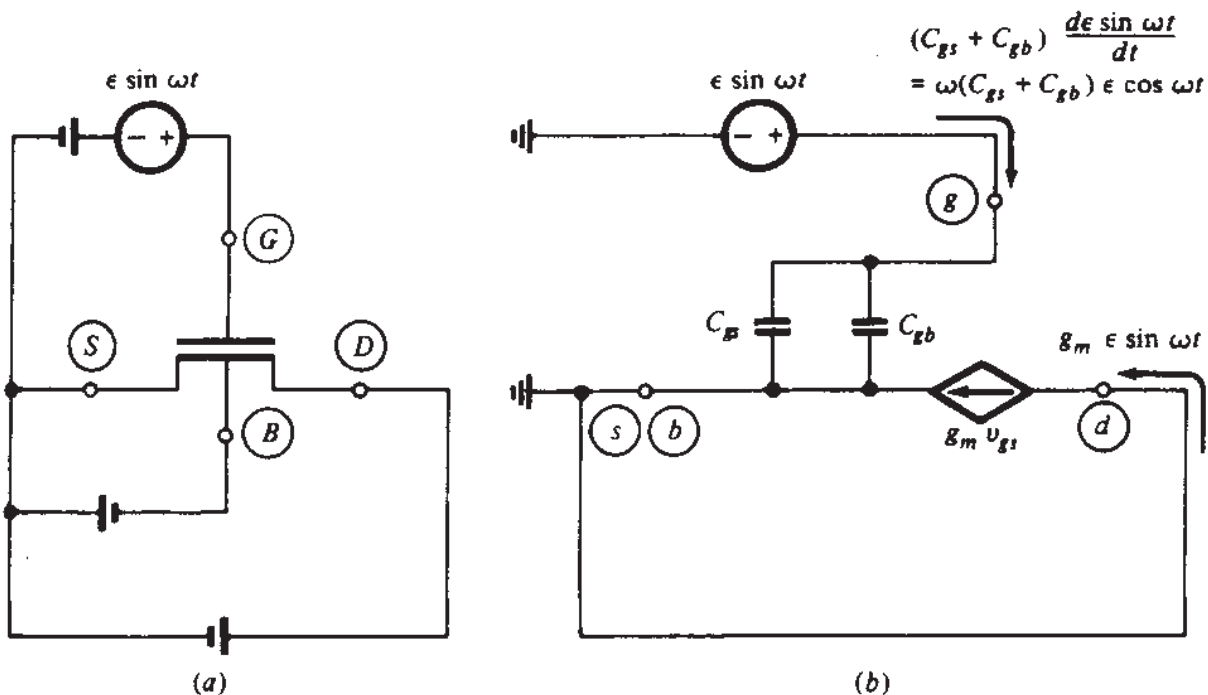


FIGURE 8.20

(a) A transistor operating in the flat part of the I_{DS} - V_{DS} characteristics with a sinusoidal small-signal excitation; (b) the small-signal equivalent circuit for (a).

the “short-circuit current gain” magnitude a_i as the ratio of the *amplitude* of the small-signal drain current to the *amplitude* of the small-signal gate current, we have

$$a_i = \frac{g_m}{\omega(C_{gs} + C_{gb})} \quad (8.3.29)$$

For $\omega \rightarrow 0$ this becomes infinite, which is to be expected since at dc there is no gate current. As ω is increased, a_i drops. The *intrinsic transition frequency*, or *intrinsic cutoff frequency*, denoted by ω_{Ti} , is defined as the value of ω at which a_i drops to the value of 1. From the above equation this value is

$$\omega_{Ti} = \frac{g_m}{C_{gs} + C_{gb}} \quad (8.3.30)$$

Assuming no velocity saturation, using (8.2.17a) and (8.3.24), and neglecting C_{gb} , we get

$$\omega_{Ti} \approx \frac{g_m}{C_{gs}} = \frac{3}{2} \frac{\mu(V_{GS} - V_T)}{\alpha L^2} = \frac{3}{2} \omega_o, \quad \text{strong inversion} \quad (8.3.31)$$

where ω_o is defined in (8.3.6). The value of ω_{Ti} is outside the region of validity of the model used, so it should initially be looked at with suspicion. However, calculation of ω_{Ti} using higher-frequency models (Chap. 9) gives essentially the same value. The reason for this is that the above calculation involves only the *magnitudes* of the currents. For these, errors using the simple model are not large at these frequencies. However, significant *phase* errors can result from using the models derived in this chapter at such high frequencies.

As an example, consider a transistor with $\mu = 600 \text{ cm}^2/(\text{V}\cdot\text{s})$, $L = 1 \text{ }\mu\text{m}$, and $\alpha = 1.2$, operating in saturation with $V_{GS} - V_T = 1 \text{ V}$. From (8.3.31) we have $\omega_{Ti} = 75 \text{ Grad/s}$. This corresponds to $f_{Ti} = \omega_{Ti}/2\pi = 11.9 \text{ GHz}$.

In practice, the current gain of a complete transistor becomes unity at some frequency less than ω_{Ti} because of the presence of extrinsic parasitic elements (Sec. 8.4). Velocity saturation can further limit the cutoff frequency; these effects are discussed in Sec. 9.6.

MODERATE INVERSION. In Sec. 7.4.3 it was mentioned that the moderate-inversion region is sometimes neglected in developing expressions for the various charges. The approach taken in such cases is to consider weak- and strong-inversion expressions as valid in adjacent regions. The resulting error is often acceptable for charge evaluation. However, such charge expressions will result in large error in predicting *capacitances*. This will be seen later in this section when we compare the above approach to results derived from complex general models valid in all regions of inversion.

As was true for small-signal conductances, to evaluate the small-signal capacitances in moderate inversion one resorts to general models, which are valid in all regions of inversion (see below). Interpolation modes are also used; an example will be shown in Table 8.1 of the end of this chapter.

WEAK INVERSION. In weak inversion things are very simple. The inversion layer charge is negligible throughout the length of the channel, and the gate “sees” the depletion region directly through the oxide. A small increase in V_B will cause some charges to enter through the substrate terminal, and this will be balanced by some gate charge leaving through the gate terminal. The corresponding value of C_{gb} can be found by using the weak inversion Q_G from (7.4.34) and (7.4.33) in (8.3.5):

$$C_{gb} = C_{ox} \frac{\gamma}{2\sqrt{\gamma^2/4 + V_{GB} - V_{FB}}} \quad (8.3.32)$$

Varying V_S or V_D in weak inversion can vary the inversion layer charge drastically in a relative sense. However, this charge remains negligible compared to Q_G and Q_B , and can play no significant role in the charge-balancing process. Hence the gate and substrate do not “feel” the variations of V_S and V_D , and the corresponding ΔQ_G and ΔQ_B in Fig. 8.15*b* and *c* are negligible. The values for the corresponding capacitances are then†

$$C_{gd} \approx C_{gs} \approx C_{bd} \approx C_{bs} \approx 0 \quad (8.3.33)$$

An intrinsic cutoff frequency can be defined for weak inversion (with $V_{DS} > 5\phi_t$) in a similar manner as in strong inversion, and Fig. 8.20 is still valid. Using (8.3.33) in (8.3.30) we get

$$\omega_{Ti} \approx \frac{g_m}{C_{gb}} \quad (8.3.34)$$

Using the weak-inversion expressions for g_m and C_{gb} , it can be shown that (Prob. 8.22)

$$\omega_{Ti} = \frac{\mu\phi_t}{L^2} \frac{I_{DS}}{I_M}, \quad \text{weak inversion} \quad (8.3.35)$$

†This discussion only considers the effect of the source and drain through the inversion layer charge. In addition to this, one has the proximity capacitances between the gate *inside* the broken line in Fig. 7.1 and the *inside* side walls of the n^+ regions.⁶⁷ These capacitances are observed in weak inversion and depletion, but reduce to zero in strong inversion, since the inversion layer then acts as a shield between the gate and the inside side walls of the n^+ regions. See “Small-Dimension Effects” below.

where I_{DS} is the current at the particular operating point and I_M is the maximum achievable current in weak inversion (i.e., the current at the upper limit of the weak inversion region).

As an example, consider a device with $\mu = 600 \text{ cm}^2/(\text{V}\cdot\text{s})$ and $L = 1 \text{ }\mu\text{m}$, operated at room temperature with I_{DS} equal to one-fifth the maximum weak inversion current. Then (8.3.35) gives $\omega_{Ti} = 311 \text{ Mrad/s}$. This corresponds to $f_{Ti} = \omega_{Ti}/2\pi = 49 \text{ MHz}$. The same device was seen earlier to have a f_T of 11.9 GHz in strong inversion, with $V_{GS} - V_T = 1 \text{ V}$.

GENERAL CHARGE SHEET MODEL. In Sec. 7.4.5, we have shown how the complete charge sheet model of Sec. 4.3 can be used to evaluate the various charges. The resulting expressions are valid in all regions of inversion. These charges can be differentiated to produce capacitances (not an easy task). The resulting expressions are too lengthy,⁵² unless an approximation is used for the bulk charge, as in Sec. 4.3.2.⁵¹⁻⁵³

To show the capacitance variation over all regions of inversion, we can fix V_{DS} and plot vs. V_{GS} . This was the approach taken in the plots of Figs. 7.9 and 8.14. The resulting capacitance plots are shown in Fig. 8.21 and agree very well with experiment in all regions. In Fig. 8.22, we show an expansion of the horizontal scale around the moderate inversion region. In this region both strong- and weak-inversion expressions fail completely. Yet, in some models for circuit CAD, weak and strong inversion expressions are assumed to be valid in adjacent regions.

In Fig. 8.23, we compare g_{mb}/g_m , C_{bs}/C_{gs} , and C_{bd}/C_{gd} . It is seen that (8.3.15) is approximately verified.

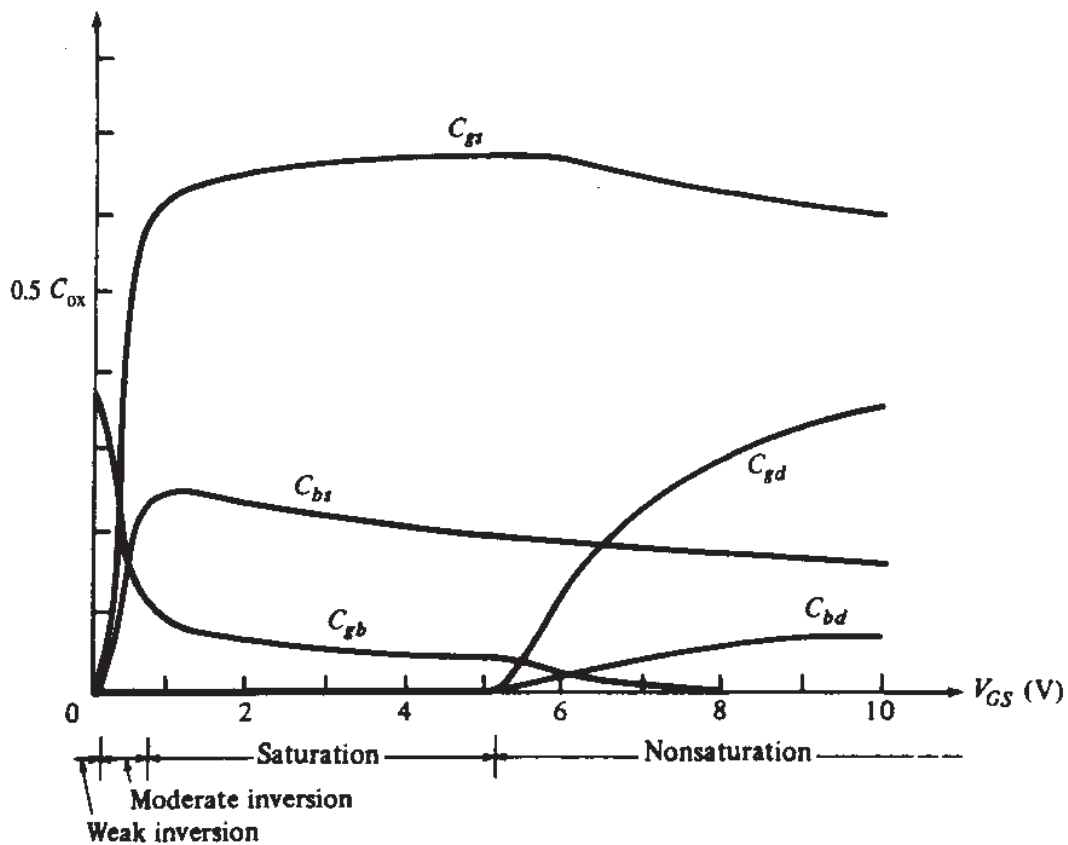
Both the simplified charge sheet model⁵¹⁻⁵³ and the EKV interpolation model⁵⁴⁻⁵⁷ result in capacitance expressions which are valid in all regions of inversion. An example is shown at the end of this chapter.

DEPLETION. In depletion, the only relevant capacitance of the intrinsic device is C_{gb} .† Using the corresponding gate charge (7.4.50) in the definition (8.3.5), we obtain the same expression as in weak inversion:

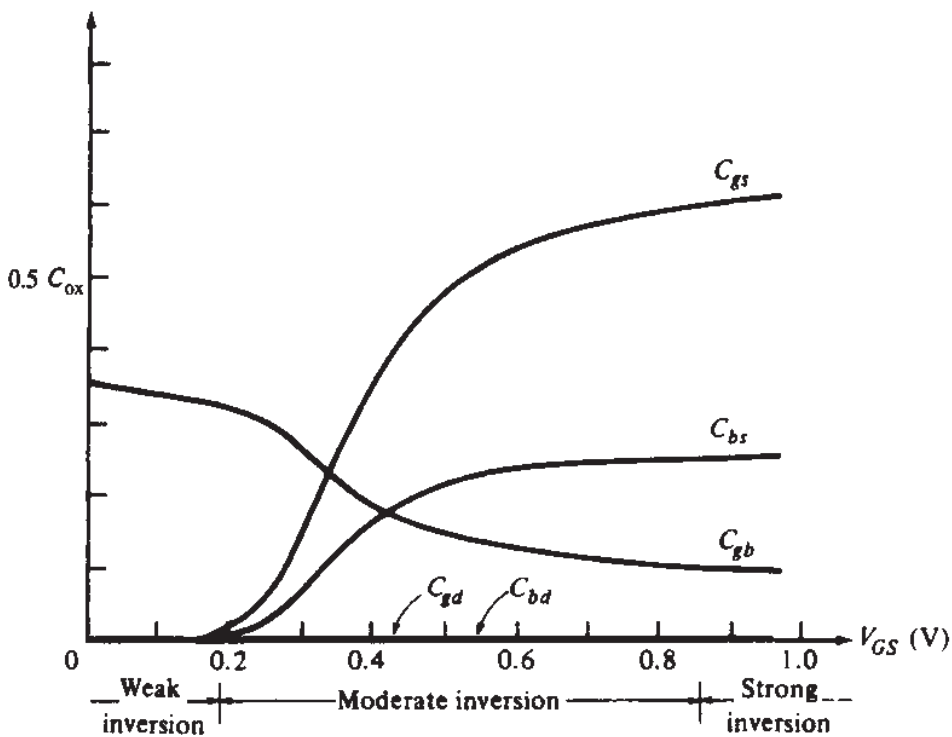
$$C_{gb} = C_{ox} \frac{\gamma}{2\sqrt{\gamma^2/4 + V_{GB} - V_{FB}}} \quad (8.3.36)$$

It should be noted that the above expression has been developed by assuming a perfect depletion region with a sharp edge. However, if V_{GB} gets close to V_{FB} (within a few ϕ_t), such a region cannot be defined clearly and the above expression will be somewhat in error. More exact calculations can be done by taking into account the distribution of carriers below the oxide (Appendix F) in conjunction with (2.6.8).

†See, however, the preceding footnote concerning the effect of the inside side walls.


FIGURE 8.21

Example of small-signal capacitances versus V_{GS} , with $V_{DS} = 4$ V and $V_{SB} = 0$, as predicted by accurate calculations.⁶²


FIGURE 8.22

Part of the plot of Fig. 8.21, expanded around the moderate inversion region.⁶²

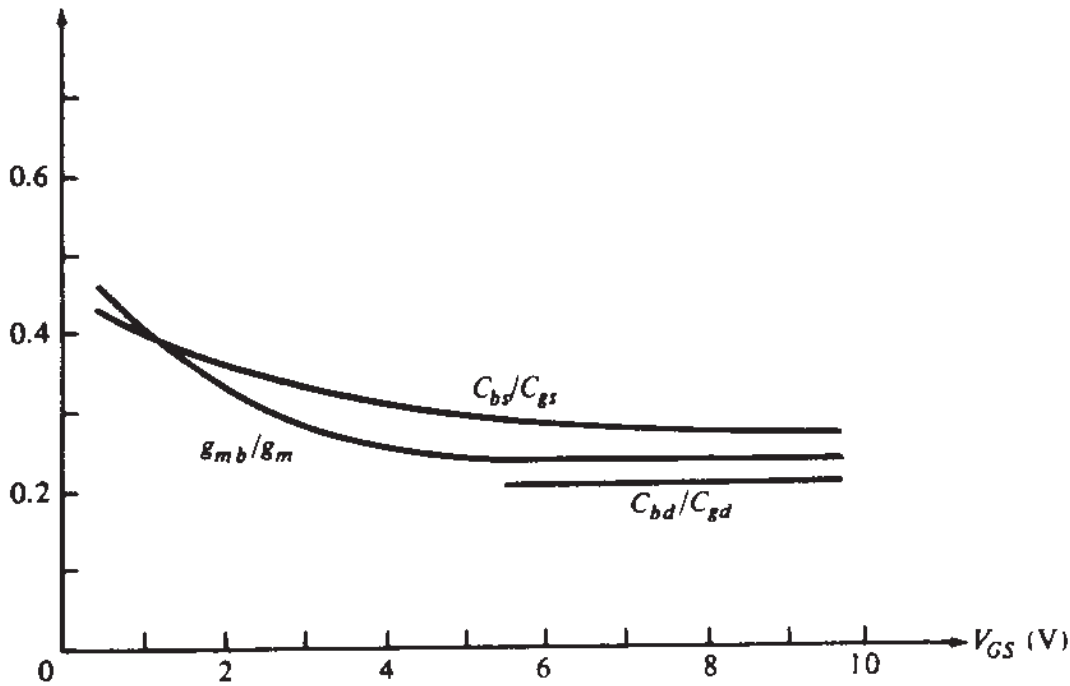


FIGURE 8.23

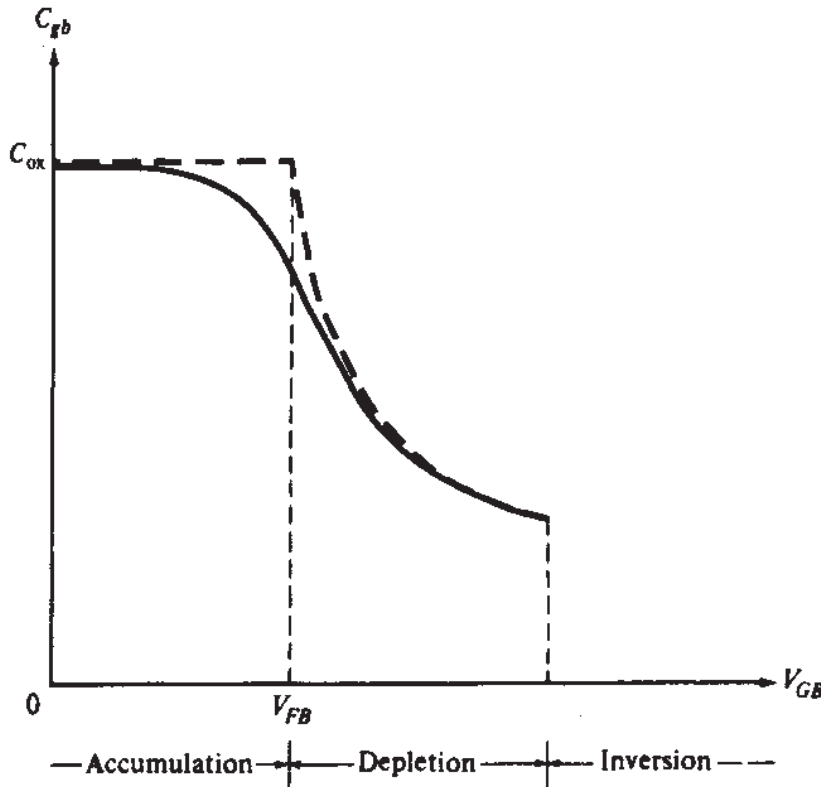
Comparison of capacitance ratios to transconductance ratios for the device of Fig. 8.21, with $V_{DS} = 4$ V and $V_{SB} = 0$, as they result from accurate calculations.⁶²

ACCUMULATION. In developing charge expressions for the accumulation region (Sec. 7.4.7), we assumed that an abundance of holes creates a conductive sheet right under the oxide. This picture or direct use of the gate charge (7.4.51) in the definition for C_{gb} (8.3.5) results in

$$C_{gb} \approx C_{ox} \quad (8.3.37)$$

The above argument about the “conductive sheet” is really valid only deep in accumulation. If V_{GB} is only slightly lower than V_{FB} , the “sheet” will not have formed satisfactorily and (8.3.37) will not be accurate. Again, C_{gb} can be calculated accurately by considering the distribution of mobile carriers with depth (Appendix F) in conjunction with (2.6.8). Figure 8.24 shows C_{gb} in the regions of accumulation and depletion. The broken line represents (8.3.36) and (8.3.37), and the solid line represents a more realistic behavior.

SMALL-DIMENSION EFFECTS. The evaluation of intrinsic capacitances of short-channel devices must take into account several effects that we could conveniently neglect in the case of long-channel devices. Such effects include velocity saturation, channel-length modulation, and of course the two-dimensional charge-sharing or DIBL effects. Some of these effects can approximately be taken into account by using “effective” values for such parameters as the threshold voltage and the gate width in the long-channel expressions, as was done in modeling the current in Chap. 6. To do a better job, the charges must be evaluated including such effects,


FIGURE 8.24

Gate-substrate capacitance in accumulation and depletion plotted versus V_{GB} . Broken line: simple model; solid line: accurate model.

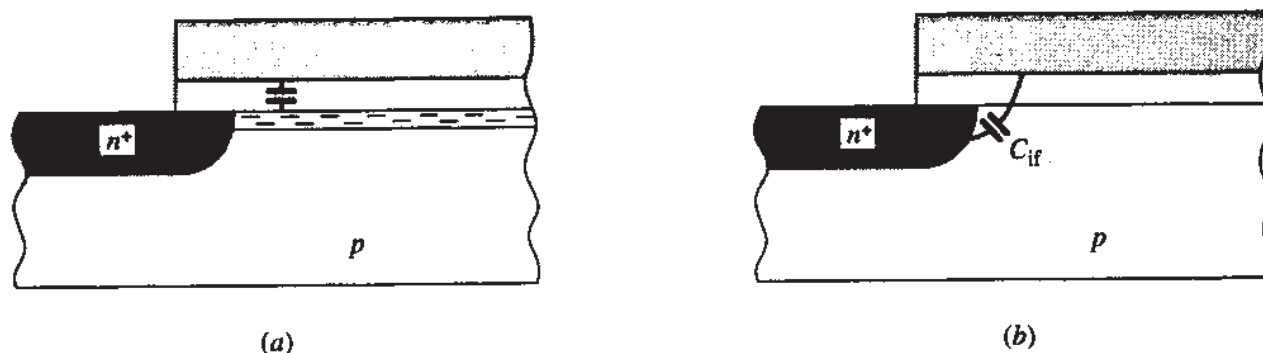
and the resulting expressions differentiated with respect to the terminal voltages, as we have done above. It should be emphasized that our definitions in (8.3.1) to (8.3.5) are still valid. However, the results are invariably complicated; the interested reader is referred to the literature.^{64-76,78,80,81,83,84,88,90,94,43}

An effect which becomes important in short-channel devices is illustrated in Fig. 8.25,⁷¹ where the situation near the source is shown. In Fig. 8.25a, a strong inversion layer is assumed formed, and a capacitance between it and the gate exists; this gives rise to the capacitance C_{gs} previously discussed. In Fig. 8.25b, the device is assumed to be off. Here the absence of an inversion layer allows the gate to “see” the inner wall of the source region directly. This gives rise to an “inner fringing” capacitance C_{if} (“outer” fringing capacitances also exist, and will be discussed in Sec. 8.4).† A similar capacitance exists at the drain end, when the inversion layer does not “screen” the gate from the inner wall of the drain, i.e., at cutoff and in saturation. As a device is driven from deep nonsaturation into saturation, the effect of the inversion

†The capacitance C_{if} has been evaluated using electrostatics, and is found to be:⁹⁵

$$C_{if} = W \frac{2}{\pi} \epsilon_s \ln \left[1 + \frac{d_j}{t_{ox}} \sin \left(\frac{2}{\pi} \frac{\epsilon_{ox}}{\epsilon_s} \right) \right]$$

where d_j is the junction depth.

**FIGURE 8.25**

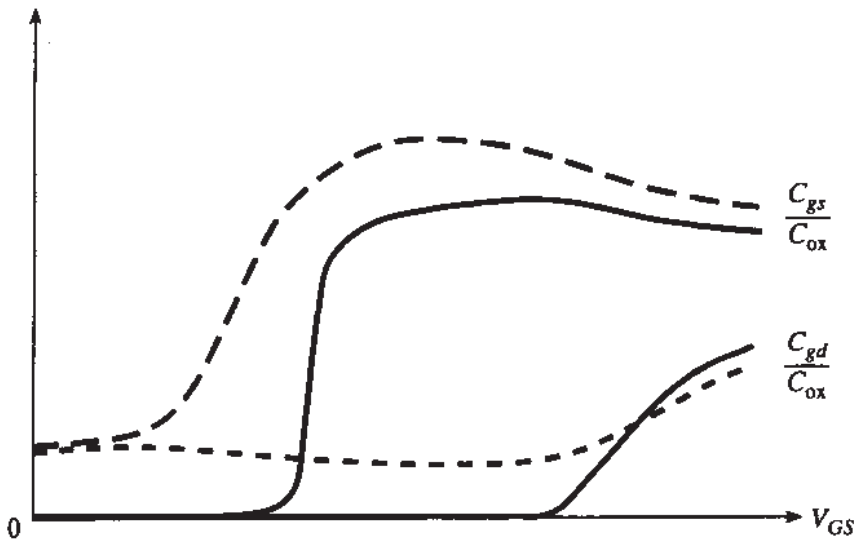
Intrinsic capacitances near the source. (a) In the presence of a strong-inversion layer; (b) in the absence of a strong inversion layer.

layer near the drain goes from fully screening the gate from the inner wall, to practically no screening at all; thus, the effective inner fringing capacitance is really a function of how far we are from saturation.[†] The inner fringing capacitance exists also in long-channel devices, but is usually small in comparison to the other capacitances.

For small-dimension devices, one typically obtains capacitance-voltage plots which are qualitatively similar to those of long-channel devices, but the details of the plots, and the actual values, can be significantly different.⁷¹⁻⁷⁶ An example is shown in Fig. 8.26, where we indicate the general behavior of the normalized gate-source and gate-drain capacitances, for a long-channel and a short-channel device, versus V_{GS} , for a fixed V_{DS} (the same type of plot as in Figs. 8.21 and 8.22). Several differences are apparent.⁷¹⁻⁷⁶ For the short-channel device, C_{gs} rises at smaller V_{GS} values, which is to be expected from the effect of DIBL on the effective threshold (Chap. 6). Before it rises, it is not zero, in contrast to the case for the long-channel device; this is because of the inner fringe capacitance discussed above. The combination of the various short-channel effects results in a larger maximum capacitance for the short-channel device, as seen. Finally, it is seen that for the short-channel device it is not easy to identify particular regions of operation on the plot, because the transition from region to region is very gradual (this trend is also apparent in the current of short-channel devices; for example, in saturation V_{DS} continues to affect the current through DIBL and channel-length modulation, and makes the transition from nonsaturation to saturation less distinct than for long-channel devices). The normalized gate-drain capacitance for the short-channel device is seen to be nonzero at low V_{GS} values, and to exhibit a very gradual transition from region to region. Very gradual transitions are also seen in C_{gb} plots. Finally, C_{gb} can be very small for short-channel devices; this may be expected from the material in Chap. 6, since a large part of the

[†]A measure of this is the ratio C_{gd}/C_{ox} which, as seen in Fig. 8.18, goes from 0.5 to 0 as V_{DS} is raised. Thus, the following empirical relation has been suggested for the effective inner-fringing capacitance:⁷⁶

$$C_{if,eff} = C_{if} \left(1 - 2 \frac{C_{gd}}{C_{ox}} \right)$$

**FIGURE 8.26**

Normalized gate-source and gate-drain intrinsic capacitances for a long-channel device (solid line) and a short-channel device (broken line), for $V_{SB} = 0$ and fixed V_{DS} .

depletion region charge is controlled by the source and drain fields, and the effect of the substrate potential on the device is weakened.

The thin oxide effects described in Sec. 6.9 can also affect the device capacitances.⁹⁶ For example, depending on device type, oxide thickness, polysilicon gate doping, and voltage polarity, a depletion region can form in the polysilicon gate immediately above the oxide, and the corresponding capacitance will be in series with the oxide capacitance. This can decrease the gate capacitance by over 10 percent.⁹⁶

The study of small-dimension effects can be greatly aided by two-dimensional computer simulations,⁷¹⁻⁷⁴ correlated with measurements. The latter can be very difficult since the intrinsic capacitances of small devices are small, and their effect can be masked by the drain current and by parasitic capacitances due to extrinsic effects, packaging, and measuring circuitry. The measurement of MOS transistor capacitances has been the subject of several studies.^{60,65,67,68,71,74,76,92,93}

8.4 SMALL-SIGNAL MODELING FOR THE EXTRINSIC PART

The “extrinsic” part of a transistor is everything outside the broken line in Fig. 8.27a. In this figure and in the top view in Fig. 8.27b we show some symbol definitions to be used in the following discussion.

The charge storage effects associated with the extrinsic part can be modeled by using six small-signal capacitances (one between each pair of terminals), as shown in Fig. 8.28. In the symbols used for these capacitances, the first two subscripts indicate the associated device terminals and the subscript *e* stands for *extrinsic*. If the transistor happens to be inside a well on a CMOS chip (Sec. 1.6), then one must consider also the capacitance due to the *pn* junction between the well and the common substrate on which the well has been formed. This capacitance is denoted by $C_{bb'}$ in

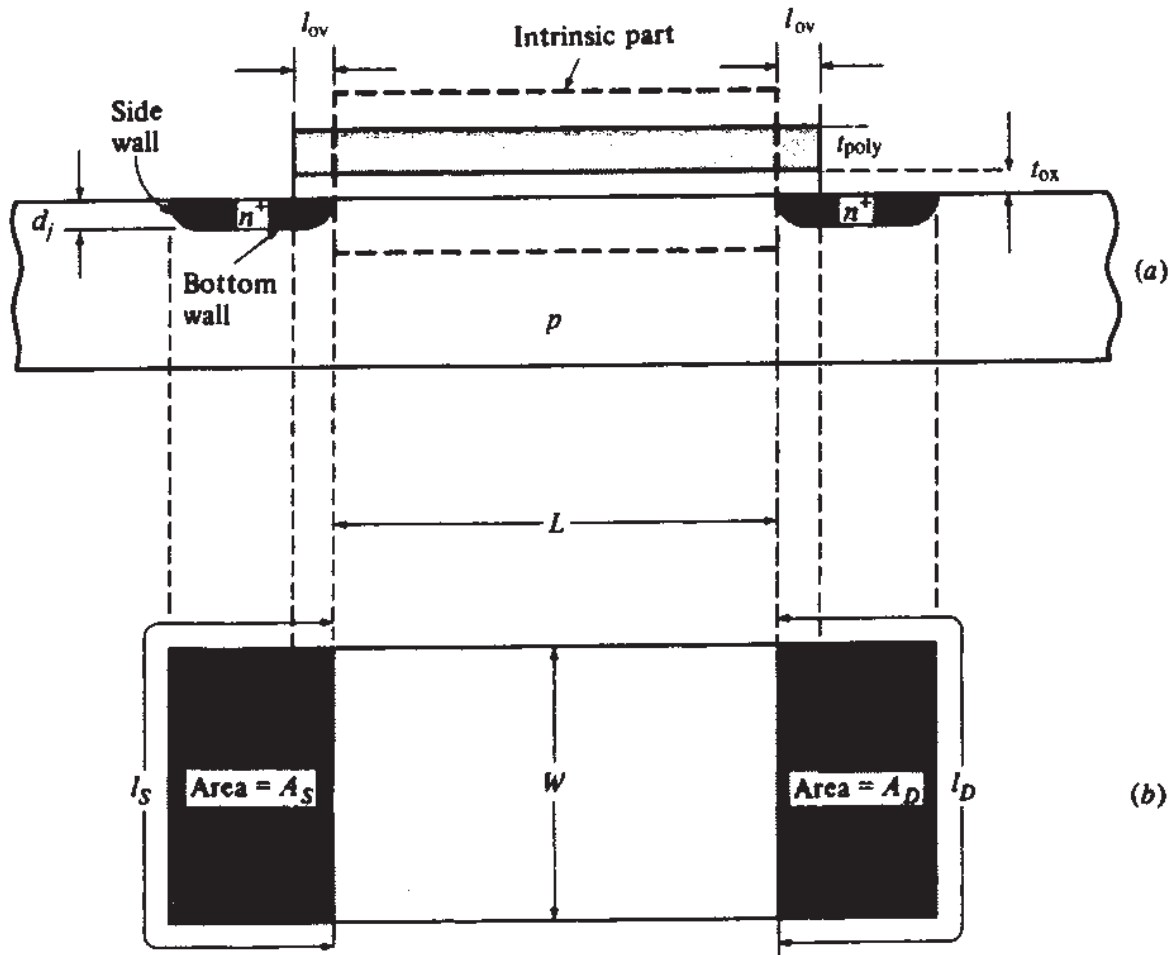


FIGURE 8.27

A MOS transistor: (a) cross section; (b) top view.

Fig. 8.28, with b corresponding to the transistor's body and b' corresponding to the common chip substrate. The dotted box represents the model of the intrinsic part, which has already been discussed in the previous two sections. We now concentrate on the seven extrinsic capacitances. Each of these can easily be associated with parts of the physical transistor structure, as discussed below.

GATE CAPACITANCES. Capacitances C_{gse} and C_{gde} are oxide capacitances. Each consists of three components, C_{ov} , C_{of} , and C_{top} , as shown in Fig. 8.29. If the doping of the source/drain diffusions is high, the overlap capacitance C_{ov} can be modeled as a nearly linear "parallel plate" capacitance with the oxide as the dielectric (provided the polysilicon depletion effect can be neglected). The component C_{of} is the outer fringing capacitance, which can be calculated from basic electrostatics.⁹⁵ All three capacitances are approximately proportional to W , and thus C_{gse} and C_{gde} can be written in the form

$$C_{gse} = C_{gde} = WC_o'' \quad (8.4.1)$$

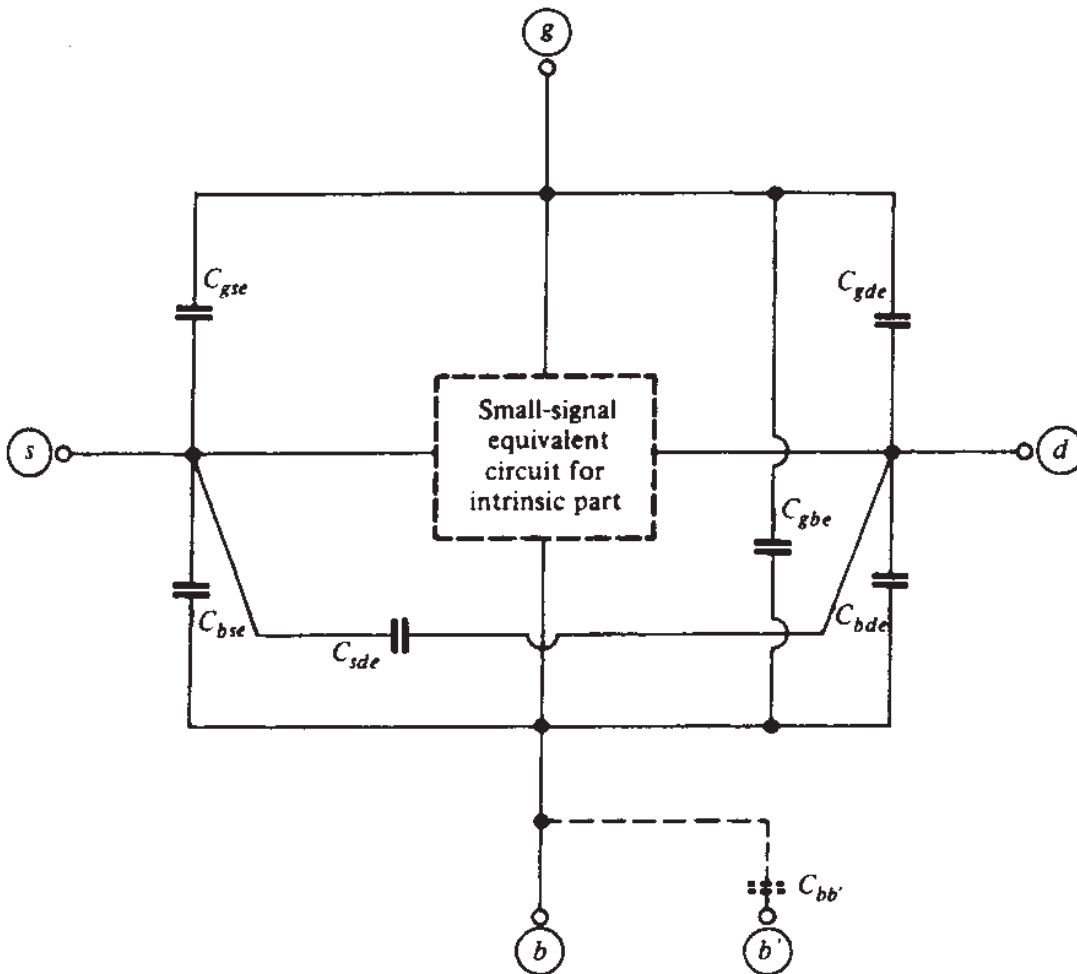


FIGURE 8.28
Extrinsic transistor capacitances added to an intrinsic small-signal model.

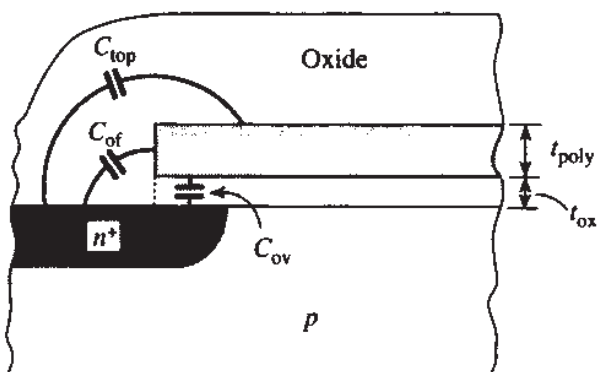


FIGURE 8.29
Extrinsic gate-diffusion capacitances.

where C_o'' is a "capacitance per unit width," including overlap, fringing, and top contributions.[†] This quantity is often estimated using 2D numerical simulations or measurements. In LDD devices, the overlap capacitances are found to be voltage-dependent, due to the light drain doping.^{86,91,97,42,43}

There also exists a parasitic capacitance C_{gbe} between gate and substrate along the channel length, outside the immediate channel area. This capacitance is proportional to the channel length, and can thus be expressed as

$$C_{gbe} = 2LC_{ob}'' \quad (8.4.2)$$

where C_{ob}'' is a "capacitance per unit length" on each side of the channel. The value of the latter is not straightforward to calculate in some technologies, as is evident from Fig. 6.13b. Numerical calculations or empirical estimates are used.

JUNCTION CAPACITANCES. The substrate-source and substrate-drain junctions give rise to small-signal capacitances C_{bse} and C_{bde} . Each of these is caused by a "bottom wall" part (over areas A_S and A_D in Fig. 8.27) and two "side wall parts," one due to the outer wall (along L_S and L_D in Fig. 8.27) and one due to the inner wall of each region (along W in Fig. 8.27). The reason one needs to consider such components is that, in general, the doping concentration of the p side for each of them is, in general, different, the details depending on device construction (see Sec. 1.6). For LOCOS processes, the capacitance per unit area along the outer wall can typically be several times that of the bottom wall, and comparable to that of the inner wall. STI (shallow-trench isolation) processes exhibit smaller outer wall capacitance per unit area than LOCOS processes.

The bottom wall part of C_{bse} and C_{bde} can be expressed in terms of a capacitance per unit area. The sidewall capacitances per unit area, on the other hand, are not constant with depth and must normally be integrated over the sidewall areas for accurate calculations. A further complication arises because the sidewalls are not "plane," but are instead nearly cylindrical. For approximate calculations, one uses effective sidewall capacitances per unit length, which only need be multiplied by the corresponding sidewall lengths to provide the total sidewall capacitances. According to the above, if A_S and C'_{js} represent the source bottom wall area and capacitance per unit area, l_S and C''_{jsf} represent the source outer sidewall length (Fig. 8.27b) and capacitance per unit length, and if W and C''_{jsc} represent the source inner sidewall length and capacitance per unit length, we will have

$$C_{bse} = A_S C'_{js} + l_S C''_{jsf} + W C''_{jsc} \quad (8.4.3)$$

[†]It can be shown⁹⁵ that C_o'' can be approximated by

$$C_o'' = l_{ov} C'_{ox} + \frac{2}{\pi} \epsilon_{ox} \ln \left(1 + \frac{t_{poly}}{l_{ox}} \right) + C''_{top}$$

where l_{ov} is the overlap distance and t_{poly} is the thickness of the poly gate, as shown in Fig. 8.29. The quantity C''_{top} can be about 10 percent of the total,⁹⁴ and increases somewhat with L .

Similarly, with A_D , C'_{jd} , l_D , C''_{jdf} , W , and C_{jdc} the corresponding quantities for the drain, we will have

$$C_{bde} = A_D C'_{jd} + l_D C''_{jdf} + W C''_{jdc} \quad (8.4.4)$$

The values for the C'_j and C''_j are usually provided for zero junction reverse bias. Under nonzero reverse bias (V_{SB} or V_{DB}), the new values of these parameters can be approximately found by assuming a functional dependence as in (1.5.22).

For devices inside a well on a CMOS chip, a third junction capacitance $C_{bb'}$ must be considered, as already explained. For this capacitance, we have

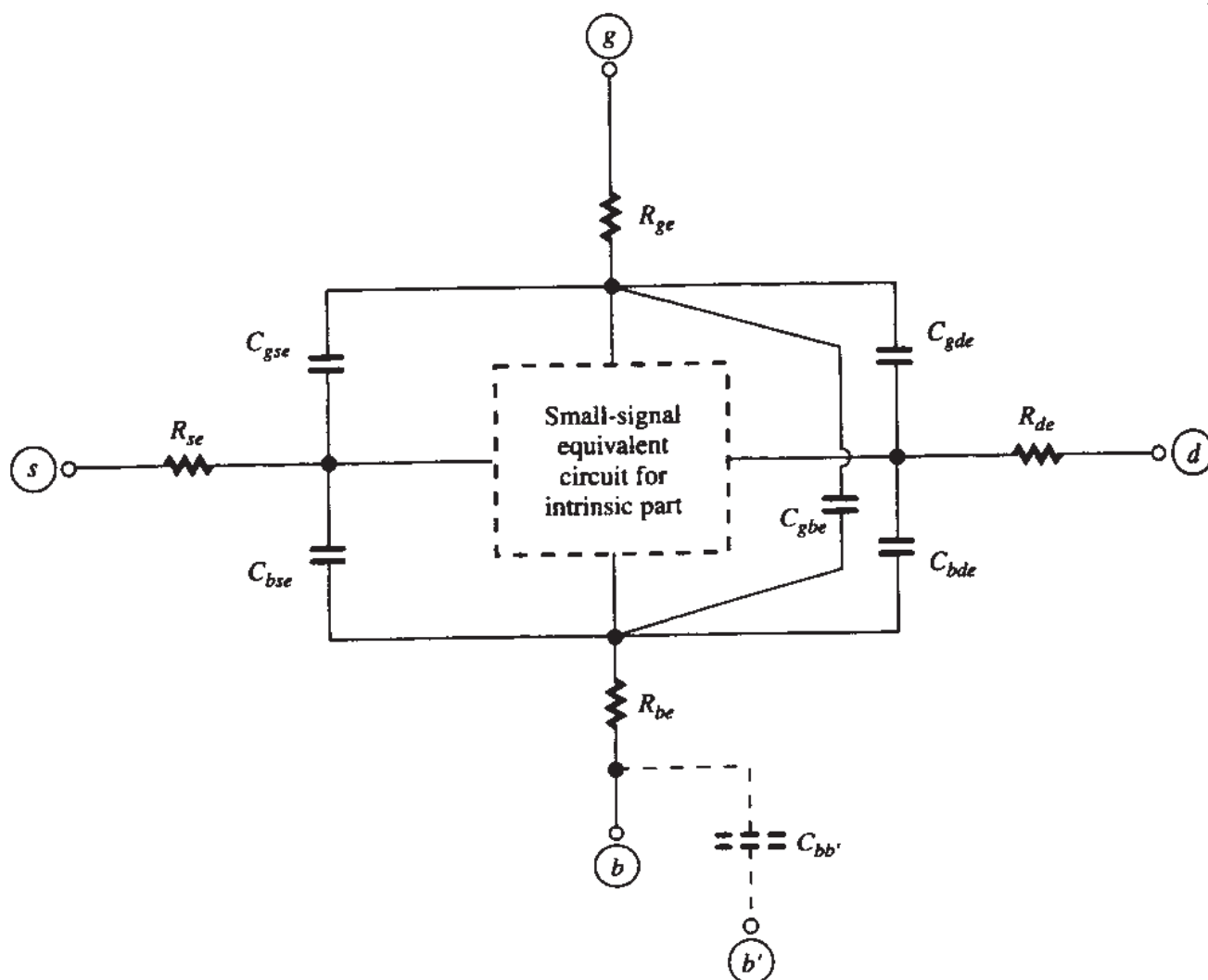
$$C_{bb'} = A_W C'_{jw} + l_W C''_{jw} \quad (8.4.5)$$

where A_W and C'_{jw} are the well's bottom wall area and capacitance per unit area, l_W is the well's sidewall length (the total perimeter as seen from above), and C''_{jw} is the well's sidewall capacitance per unit length. It is obvious that, if a group of more than one transistor share the same well, $C_{bb'}$ must be included only once in the model for the group.

SOURCE-DRAIN PROXIMITY CAPACITANCE. As with any objects in close proximity, a capacitance exists between the source and drain n^+ regions, denoted by C_{sde} in Fig. 8.28. Because of the complicated shapes involved, the value of this capacitance is difficult to evaluate. However, it is, in general, very small and can be neglected in comparison to other capacitances, unless the channel is very short.

RESISTANCES. The above capacitances are very easy to incorporate in a complete small-signal model at relatively low frequencies, since they appear in parallel with the corresponding intrinsic capacitances in the model of Fig. 8.17 (compare to Fig. 8.28). However, a more complete model for the extrinsic part should include parasitic resistances as well. Such resistances are those of the source and drain regions and their contacts, and of the substrate and the gate material and their contacts (see Sec. 6.8). The modeling of these resistors is complicated,^{98-102,42} especially in LDD devices (Sec. 6.6), in which the resistances can show significant voltage dependence.^{88,42,97} At high frequencies, the magnitude of the impedance of the capacitances drops to the point where it becomes comparable to the resistances; hence, the latter should be taken into account. This is done most simply by approximating the resistive paths by a few lumped elements. An example is shown in Fig. 8.30, where the resistance symbols with subscripts s , d , g , and b model the resistive materials of the source region, drain region, gate, and substrate, respectively. Unfortunately, no general rules can be given for the frequency where such effects of these resistances become important. This depends greatly on the resistivity of the regions, their geometry, the way they are contacted, etc.

A cutoff frequency can be defined for the complete transistor as the frequency at which the small-signal current gain becomes unity, in analogy to the intrinsic cutoff frequency defined earlier in conjunction with Fig. 8.20a. Because of the presence of the extrinsic elements, the cutoff frequency for the complete transistor is smaller than for the intrinsic part (Sec. 9.6).

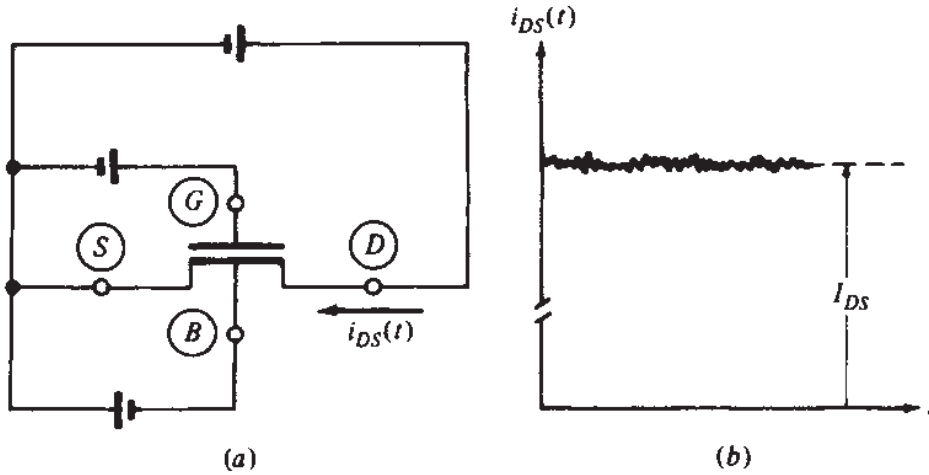
**FIGURE 8.30**

Small-signal equivalent circuit of the MOS transistor, including extrinsic capacitances and resistances.

8.5 NOISE

8.5.1 Introduction

It has so far been assumed that the drain current of a MOS transistor varies with time only if one or more of the terminal voltages vary with time. This is not exactly true. A careful examination of the current reveals minute fluctuations, referred to as *noise*, which are present whether externally applied signals are present or not. Such fluctuations can occur due to several mechanisms (see below). Noise can interfere with weak signals when the transistor is part of an analog circuit,^{103–105} so ways to predict and possibly reduce noise are very important. For this reason, the subject of noise in MOS transistors has received extensive treatment in the literature. This section is devoted to this subject. We have chosen to treat noise as part of our discussion of small-signal modeling, because noise is, in a sense, an internally generated small signal in the device, and can be modeled with appropriate additions to the small-signal equivalent circuits we have already developed in this chapter.

**FIGURE 8.31**

(a) A MOS transistor biased with fixed noiseless terminal voltages; (b) the drain-to-source current for the connection in (a), including noise.

Consider a transistor with dc bias voltages, as shown in Fig. 8.31a. The total drain current, shown in Fig. 8.31b, can be expressed as

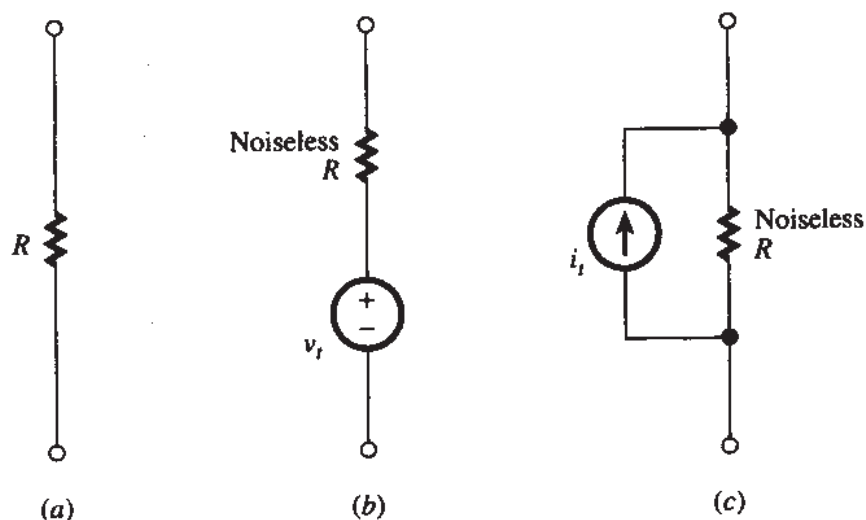
$$i_{DS}(t) = I_{DS} + i_n(t) \quad (8.5.1)$$

where I_{DS} is the ideal (bias) current and $i_n(t)$ is the noise component, which has zero average value. The instantaneous value of i_n at a given t is, of course, unpredictable. Instead, one talks about certain measures characterizing the behavior of $i_n(t)$. In noise work, such measures are the *mean square* value, denoted by i_n^2 , and the *root mean square* (rms) value, $\sqrt{i_n^2}$.

In measuring noise quantities, the amount of noise depends on the bandwidth of the measuring instrument. A common measurement involves a *very narrow* bandwidth, centered on a frequency f . The current noise spectral components within this bandwidth have a certain mean square value. The *ratio* of this value to the bandwidth, as the latter is allowed to approach zero, tends to what is called the *power spectral density*† of the current noise, denoted by $S_i(f)$. This quantity has units of square amperes per hertz. Often the square root of the power spectral density is used instead, given in $A/\sqrt{\text{Hz}}$. For a noise voltage v_n , one can similarly define a power spectral density $S_v(f)$ (in square volts per hertz) or its square root (in $V/\sqrt{\text{Hz}}$).

The total mean square noise current within an arbitrary bandwidth extending from $f = f_1$ to $f = f_2$ can be found by summing the mean square values of the indi-

†More rigorous definitions can be found in specialized texts.^{103,104}

**FIGURE 8.32**

(a) A real (noisy) resistor; (b) Thevenin noise equivalent circuit for (a); (c) Norton noise equivalent circuit for (a).

vidual components within each subbandwidth Δf . More precisely, using the power spectral density concept, we have

$$\overline{i_n^2} = \int_{f_1}^{f_2} S_i(f) df \quad (8.5.2)$$

A similar result can be obtained for voltage noise.

A well-known example of device noise is the *thermal noise* in a resistor (also called *Johnson noise* or *Nyquist noise*). It is due to the random thermal motion of the carriers in it.^{103,104} A real (noisy) resistor (Fig. 8.32a) can be represented as an ideal (noiseless) resistor R in a series with a noise voltage source as shown in Fig. 8.32b. Using concepts from statistical physics, it can be shown that the corresponding power spectral density is^{103,104}

$$S_{vt} = 4kTR \quad (8.5.3)$$

where k is Boltzmann's constant (1.38×10^{-23} V·C/K) and T is the absolute temperature. Thermal noise is said to be *white* noise, because its power spectral density, as given by the above formula, is "flat" up to extremely high frequencies (over 10^{12} Hz).

The circuit in Fig. 8.32b is a "Thevenin equivalent circuit."^{105a} This can be converted to a "Norton equivalent circuit,"^{105a} shown in Fig. 8.32c, with a noise current $i_t = v_t/R$, or equivalently with $\overline{i_t^2} = \overline{v_t^2}/R^2$; thus, the power spectral density of this noise current is

$$S_{it} = 4kT \frac{1}{R} \quad (8.5.4)$$

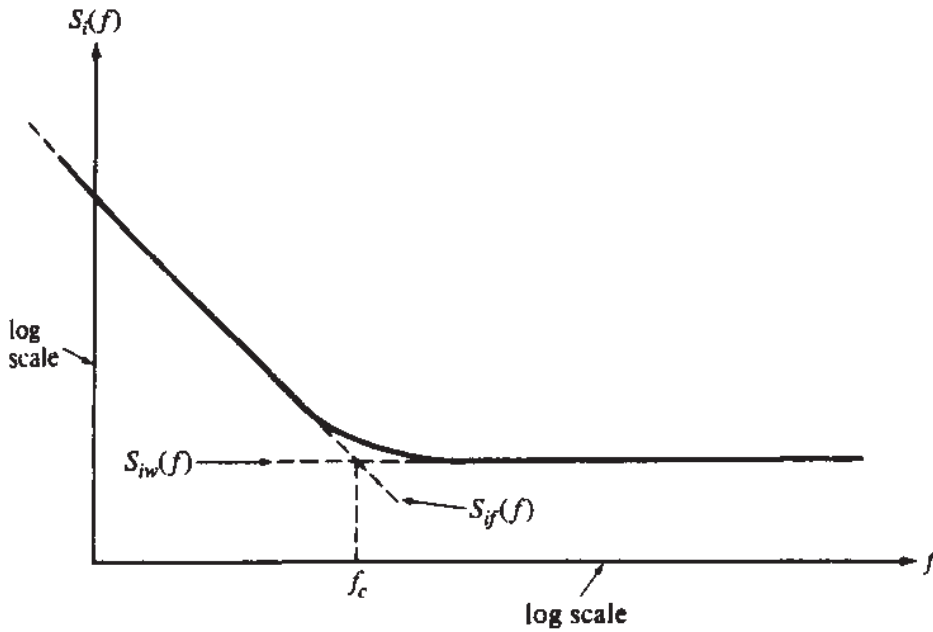


FIGURE 8.33

A typical plot of the drain-noise current power-spectral density versus frequency in log-log axes. Subscript i refer to total noise, i_w to white noise, and i_f to flicker noise.

A typical plot of power spectral density for the drain current noise of a MOS device is shown in Fig. 8.33 on log-log axes. Two distinct frequency regions, with different noise behavior in each, can be identified. These regions can be thought of as separated by a *corner frequency* f_c . Values from several hertz to several megahertz are common for this quantity depending on device construction, geometry, and bias. The type of noise dominating at high frequencies in Fig. 8.33 is *white noise* (its power spectral density is constant up to extremely high frequencies). The corresponding noise current component will be denoted by i_w . The noise dominating at low frequencies is called *flicker noise* or $1/f$ noise; the power spectral density for the current caused by this type of noise is nearly proportional to $1/f$ (Sec. 8.5.3). The current component due to flicker noise will be denoted by i_f .

The noise currents i_w and i_f are independent. In calculating the total noise mean square value due to both, one can consider the effect of each separately and then *add the individual mean square values*. To see why, consider a noise current $i_n(t)$ which consists of the two independent contributions $i_w(t)$ and $i_f(t)$. We have $i_n(t) = i_w(t) + i_f(t)$, which gives $i_n^2(t) = i_w^2(t) + i_f^2(t) + 2i_w(t)i_f(t)$. The average of $i_n^2(t)$ is $\overline{i_n^2(t)} = \overline{i_w^2(t)} + \overline{i_f^2(t)} + \overline{2i_w(t)i_f(t)}$, where bars denote averages. However, if two contributions are independent, and at least one of them has zero mean, their product averages to zero^{103,104} and thus

$$\overline{i_n^2} = \overline{i_w^2} + \overline{i_f^2} \quad (8.5.5)$$

Let now $\overline{i_n^2}$, $\overline{i_w^2}$, and $\overline{i_f^2}$ be the mean square values of the corresponding noise spectral components within a very small bandwidth, and divide both sides of (8.5.5)

by that bandwidth; allowing the bandwidth to approach zero, and recalling the concept of power spectral density discussed above, we see that $S_i(f)$, the power spectral density of the total noise, will be given by

$$S_i(f) = S_{iw}(f) + S_{if}(f) \quad (8.5.6)$$

where $S_{iw}(f)$ and $S_{if}(f)$ are the power spectral densities of the white and flicker noise components, respectively, indicated in Fig. 8.33.

We now consider MOSFET white noise and flicker noise separately.^{103-105,106-233}

8.5.2 White Noise

STRONG INVERSION. In the strong-inversion region, the white noise is actually thermal noise (also called *Johnson noise* or *Nyquist noise*) and is certainly the type of noise best characterized for the MOS transistor (most of Refs. 103, 104, 106-159 deal with this type of noise). The term *thermal* is due to the origin of this noise, which can be traced to the random thermal motion of carriers in the channel. Before embarking on an evaluation of this noise, we give here two conventional relations derived for the strongly inverted MOS transistor in Sec. 4.5.1. The drain current (assumed noiseless) was shown there to be, in the absence of velocity saturation,

$$I_{DS} = -\mu W Q'_I(V_{CB}(x)) \frac{dV_{CB}(x)}{dx} \quad (8.5.7)$$

where x is the position along the channel, $V_{CB}(x)$ is the "effective reverse bias" of the strongly inverted channel with respect to the substrate at position x (see Fig. 8.34), Q'_I is the inversion layer charge per unit area, μ is the mobility, and W is the width of the channel. Integrating this equation was seen to give, assuming a constant mobility,

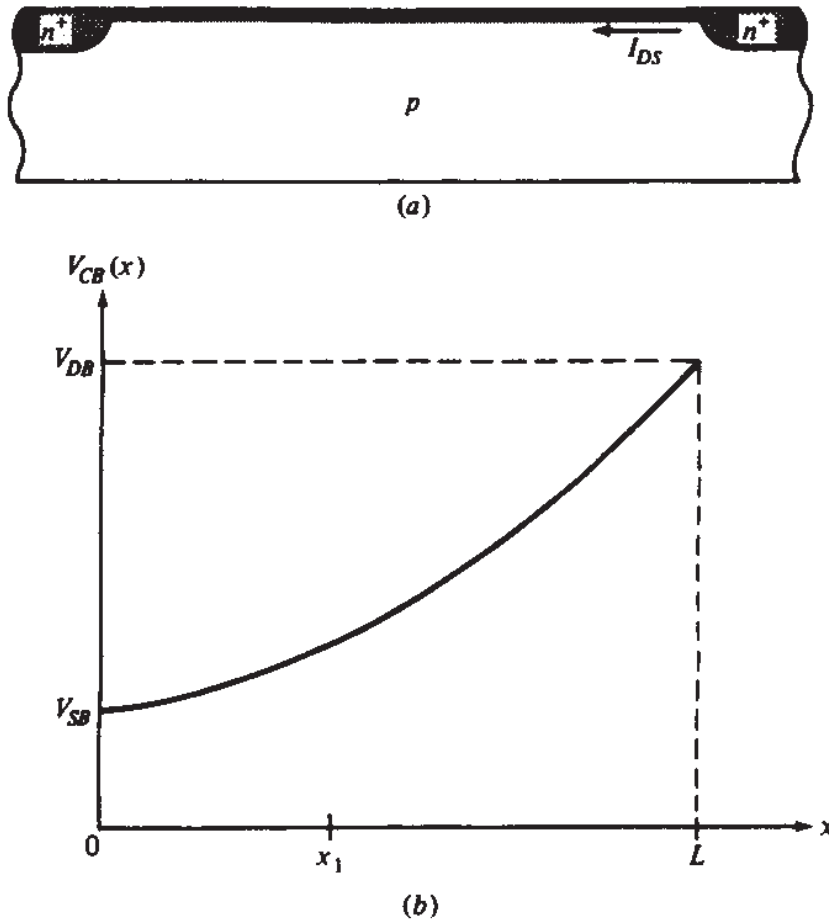
$$I_{DS} = -\frac{W}{L} \mu \int_{V_{SB}}^{V_{DB}} Q'_I(V_{CB}(x)) dV_{CB}(x) \quad (8.5.8)$$

with L being the channel length. Depending on the expression for $Q'_I(V_{CB})$, we have seen that models of differing complexity and accuracy are obtained.

In the following development, we also need an expression for the resistance ΔR of a small element of the channel of length Δx centered around a point $x = x_1$. Since $\Delta V_{CB} = I_{DS} \Delta R$ we have, from (8.5.7),

$$\Delta R = \frac{\Delta x}{-\mu W Q'_I(V_{CB}(x_1))} \quad (8.5.9)$$

where the right-hand side is, of course, positive, since $Q'_I < 0$.

**FIGURE 8.34**

(a) Part of a MOS transistor in strong inversion; (b) effective reverse bias V_{CB} vs. distance along the channel.

Assuming that the small element of the channel acts as a resistor of resistance ΔR , we will observe across it a small noise voltage Δv_i with a power spectral density $4kT \Delta R$ [see (8.5.3)]. Thus, the mean square value of Δv_i over a bandwidth B will be†

$$\overline{(\Delta v_i)^2} = 4kT \Delta R B \quad (8.5.10)$$

from which, using (8.5.9), we get

$$\overline{(\Delta v_i)^2} = \frac{4kT \Delta x}{-\mu W Q'_i(V_{CB}(x))} B \quad (8.5.11)$$

†Here B does not have to be small in order to write (8.5.10), since the power spectral density is constant (the noise is white). If the power spectral density were a function of frequency, B would have to be chosen very small.

This noise will cause noise in the drain current. To study the mechanism by which this happens, we consider the following "thought" experiment. A fictitious dc voltage source of negligible length and of very small magnitude Δv is inserted at point x_1 in the channel, as shown in Fig. 8.35a. This will create a jump Δv in the potential $V_{CB}(x)$, as shown in Fig. 8.35b,¹⁰⁴ and will cause a change in the drain current, in comparison to that in Fig. 8.34. To get a feeling for this effect, one can consider Fig. 8.36.¹³ Here we have two transistors of lengths x_1 and $L - x_1$. The dimensions of the source-drain regions connected to Δv are assumed to have shrunk to zero. If $\Delta v = 0$, the connection of the two transistors is equivalent to the single transistor in Fig. 8.34. If $\Delta v \neq 0$, the drain bias of the transistor on the left and the source bias of the transistor on the right will be disturbed. A new value for the current will then be established, along with a new potential distribution. These will correspond to the situation shown in Fig. 8.36. Let the new drain current value be $I_{DS} + \Delta i$, as noted in Figs. 8.35 and 8.36. We can write equations similar to (8.5.8) for the left and the right transistors. These will be, correspondingly,

$$I_{DS} + \Delta i = -\frac{W}{x_1} \mu \int_{V_{SB}}^{V_1} Q'_I(V_{CB}(x)) dV_{CB}(x) \quad (8.5.12)$$

$$I_{DS} + \Delta i = -\frac{W}{L - x_1} \mu \int_{V_1 + \Delta v}^{V_{DB}} Q'_I(V_{CB}(x)) dV_{CB}(x) \quad (8.5.13)$$

where V_1 is defined in Fig. 8.35b. Eliminating x_1 among these equations and using the assumption that Δv is very small easily gives (Prob. 8.24)

$$I_{DS} + \Delta i = -\frac{W}{L} \mu \int_{V_{SB}}^{V_{DB}} Q'_I(V_{CB}(x)) dV_{CB}(x) + \frac{W}{L} \mu Q'_I(V_{CB}(x_1)) \Delta v \quad (8.5.14)$$

As Δv goes to zero, $V_{CB}(x_1)$ has a well-defined value, in fact the same as in Fig. 8.34b. Recognizing in the above equation the first term on the right-hand side as I_{DS} from (8.5.8), we obtain for the change Δi :

$$\Delta i = \frac{W}{L} \mu Q'_I(V_{CB}(x_1)) \Delta v \quad (8.5.15)$$

In the above development, we have assumed that Δv is a dc voltage. However, the result will be valid even if Δv is varying with time as long as the variation is slow enough so that quasi-static behavior is maintained (Sec. 7.2). This implies frequencies several times less than ω_ϕ , just as was the case for the small-signal model we have already developed in Sec. 8.3. Let us now remove the battery and instead consider the thermal noise voltage generated across the small element of the channel centered at x_1 . Let Δv_t represent that part of the total thermal noise voltage which has frequency components in a bandwidth B within the above frequency range.[†] If Δi_t ,

[†]This assumption will result in models sufficient for many applications. Noise at higher frequencies is discussed in Chap. 9.

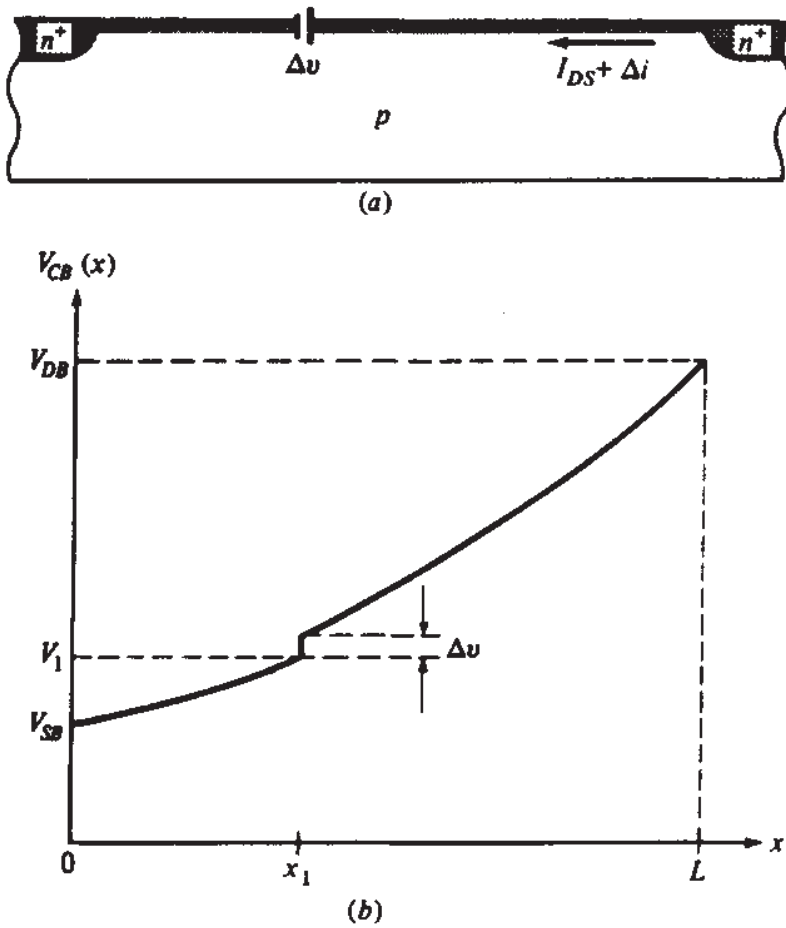


FIGURE 8.35
 (a) The transistor of Fig. 8.34a, with a fictitious dc source placed at point $x = x_1$ in the channel; (b) resulting effective reverse bias V_{CB} versus distance.

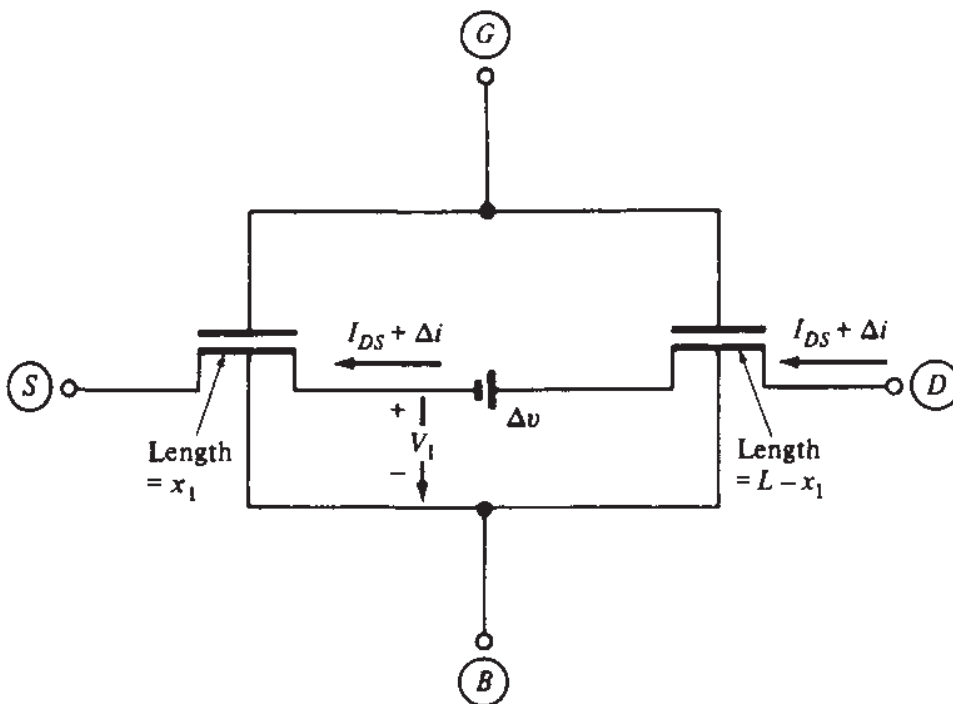


FIGURE 8.36
 Schematic representation of the situation in Fig. 8.35a.

represents the corresponding drain current variation, we will have, in analogy with (8.5.15),

$$\Delta i_t(t) = \frac{W}{L} \mu Q'_t(V_{CB}(x_1)) \Delta v_t(t) \quad (8.5.16)$$

The mean square value of Δi_t can be found by noting that, for negligibly small Δv_t , $V_{CB}(x_1)$ has a practically constant, well-defined value, in fact the same value as in Fig. 8.34b, as already mentioned. (We do not need the actual value, as will be seen. We only need to know that it is practically independent of Δv_t for negligibly small Δv_t .) Thus, the mean square value of Δi_t will be

$$\overline{(\Delta i_t)^2} = \left[\frac{W}{L} \mu Q'_t(V_{CB}(x_1)) \right]^2 \overline{(\Delta v_t)^2} \quad (8.5.17)$$

Using now (8.5.11) in the above relation, we obtain

$$\overline{(\Delta i_t)^2} = -4kT \frac{\mu}{L^2} W Q'_t(V_{CB}(x_1)) \Delta x \cdot B \quad (8.5.18)$$

This gives the contribution of the element at x_1 to the drain current noise. The contributions of all similar elements in the channel are assumed uncorrelated and with zero average value, and one can thus find the mean square value of their combined effect by adding the individual mean square values. In the limit, letting Δx become a differential and integrating over the channel length, we obtain

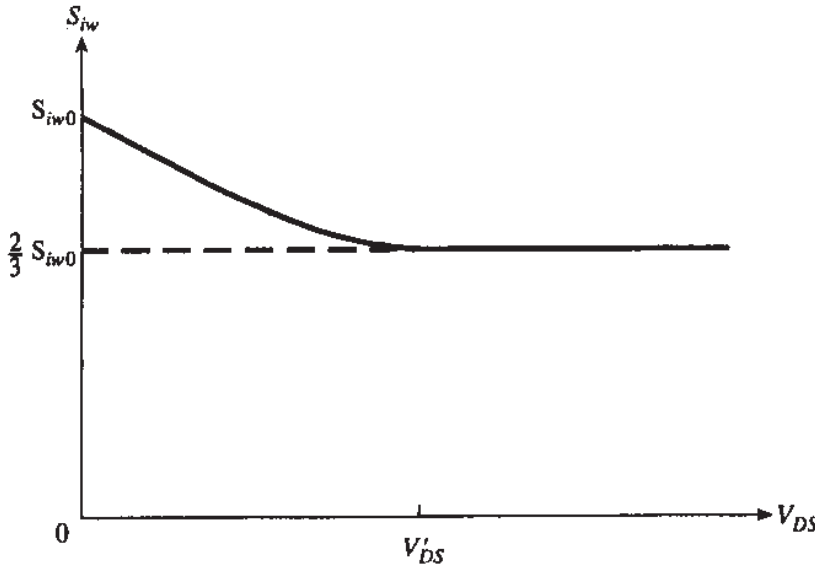
$$\overline{i_t^2} = -4kT \frac{\mu}{L^2} \left(\int_0^L Q'_t W dx \right) B \quad (8.5.19)$$

where we have used $\overline{i_t^2}$ to denote the total noise current mean square value in a bandwidth B . Recognizing the integral in the above equation as the total inversion layer charge Q_t , and dividing both sides by B , we obtain the power spectral density of the white noise

$$S_{iw} = 4kT \frac{\mu}{L^2} (-Q_t) \quad (8.5.20)$$

This equation is valid for any model, provided the appropriate expression is used for Q_t . In particular, for the *approximate* strong inversion model, we can use Q_t from (7.4.14). This gives

$$S_{iw} = 4kT \left[\frac{W}{L} \mu C'_{ox} (V_{GS} - V_T) \frac{2}{3} \frac{1 + \eta + \eta^2}{1 + \eta} \right] \quad (8.5.21)$$

**FIGURE 8.37**

Normalized power spectral density of MOS transistor thermal noise current versus V_{DS} . Operation is in strong inversion. S_{iw0} is the value of the power spectral density at $V_{DS} = 0$.

where η has been defined in (4.5.38). In nonsaturation with $V_{DS} = 0$ we have $\eta = 1$, giving

$$S_{iw} = 4kT \left[\frac{W}{L} \mu C'_{ox} (V_{GS} - V_T) \right], \quad V_{DS} = 0 \quad (8.5.22)$$

Using (8.2.22), we can see that S_{iw} is precisely $4kTg_{sd}$, where g_{sd} is the source-drain small-signal conductance. This is fully consistent with (8.5.4) and shows that, at this point, the channel behaves as a resistor of value $R = 1/g_{sd}$.

In saturation $\eta = 0$, giving

$$S_{iw} = 4kT \left[\frac{2}{3} \frac{W}{L} \mu C'_{ox} (V_{GS} - V_T) \right], \quad V_{DS} > V'_{DS} \quad \dagger \quad (8.5.23)$$

S_{iw} is plotted versus V_{DS} in Fig. 8.37. Notice that, for a given V_{GS} and V_{SB} , the noise is maximum at $V_{DS} = 0$.

As follows from the above, for a given bias the power spectral density is *independent* of frequency, at least in the range of frequencies where the assumption of quasi-static behavior is valid.

†If in (8.5.23) we use (8.2.16b) and (4.5.34), we obtain $S_{iw} = 4kT \frac{2}{3} (\alpha g_m)$ in the saturation region. In circuits literature and circuit simulators, this formula is sometimes used indiscriminately in both saturation and nonsaturation, which can lead to very wrong results. For example, at $V_{DS} = 0$, we have $g_m = 0$, which would predict zero noise if this formula were used! This is, of course, physically impossible and clearly false, as follows from (8.5.21).

A common representation of noise involves the so-called *equivalent input noise voltage*. This quantity is defined as the noise needed in the voltage between the gate and source of a hypothetical *noiseless* transistor, to produce the correct amount of noise current. Let us denote by $v_{n,eq}$ the equivalent input noise voltage. Recalling the definition of the gate transconductance, $v_{n,eq}$ must be such as to give $i_t = g_m v_{n,eq}$, and thus $i_t^2 = g_m^2 v_{n,eq}^2$. Considering the spectral components within a bandwidth B , and dividing by B , we obtain the relation between the corresponding power spectral densities:

$$\boxed{S_{vw} = \frac{S_{iw}}{g_m^2}} \quad (8.5.24)$$

where S_{vw} is the power spectral density of the equivalent input noise voltage. In analog applications, an "input" signal is intentionally superimposed on the bias between gate and source. This signal can be viewed as being in series with $v_{n,eq}$, and thus the two can be compared to discuss the resulting signal-to-noise ratio. The quantity S_{vw} has the problem that it becomes infinite at $V_{DS} = 0$, since g_m in (8.5.24) becomes zero at that point. This, of course, does not imply infinite noise in the channel; the product $g_m^2 S_{vw}$ is finite, and gives the correct value for S_{iw} .

Yet another common description of noise involves the concept of *equivalent input noise resistance*. This is a fictitious resistance of such value that its thermal noise power spectral density is S_{vw} . Since the power spectral density of the thermal noise voltage across a resistance R is given by (8.5.3), it is seen that the relation between S_{vw} and R_n is

$$S_{vw} = 4kTR_n \quad (8.5.25)$$

At $V_{DS} = 0$, R_n becomes infinite, as expected from the above discussion.

The use of the concepts of S_{vw} and R_n requires some care. Some problems that can arise in this respect are discussed in Prob. 8.26.

WEAK INVERSION. Transistors operating in weak inversion exhibit a "flat" part in the power spectral density for the noise current, just as devices in strong inversion do (Fig. 8.33). In some treatments, this is taken to be caused by thermal noise, just as in strong inversion^{139,141} (the alternate assumption¹³⁶ of *shot* noise is discussed later). Expressions for the power spectral density are then derived by taking (8.5.20) to be valid in weak inversion,¹⁴¹ on the basis of a proof given elsewhere.¹¹⁰ Q_I can be obtained from (7.4.36), repeated here:

$$Q_I = WL \frac{Q'_{I0} + Q'_{IL}}{2} \quad (8.5.26)$$

Expressions for the inversion layer charge per unit area at the source and drain ends of the channel were developed in Sec. 4.6. Using these in the above equation, and using the current expression (4.6.12) in the result, we easily get (Prob. 8.27)

$$Q_I = \frac{L^2}{2\mu\phi_t} I'_{DS} (1 + e^{-V_{DS}/\phi_t}) \quad (8.5.27)$$

where I'_{DS} is the current in the flat part of the I_{DS} - V_{DS} curve ($V_{DS} > 5\phi_t$). Using this in (8.5.20) (which is still considered valid¹⁴¹) and recalling that $\phi_t = kT/q$, where q is the electron charge, gives:

$$S_{iw} = 2qI'_{DS} (1 + e^{-V_{DS}/\phi_t}) \quad (8.5.28)$$

For large V_{DS} ($> 5\phi_t$), this reduces to $2qI_{DS}$. Although we have assumed the presence of *thermal* noise in deriving the above result, this is precisely the same value as what one would get if a different type of noise, called *shot* noise, were assumed.¹³⁶ Shot noise is associated with dc flow produced by carriers crossing a potential barrier (such as the one from source to channel), and is due to the discreteness of the arriving charges. Shot noise can be shown to have a power spectral density of $2qI$, where I is a dc current.^{123,104} Thus, whether one assumes thermal or shot noise in weak inversion, the same result is obtained.† This has led to some controversy as to which of the two types of noise is actually present in weak inversion.^{136,139,141}

An equivalent input noise voltage can be defined as before by using (8.5.24). An equivalent input noise resistance can be defined using (8.5.25).

MODERATE INVERSION AND GENERAL MODELS. As has been seen, (8.5.20) can be used in both strong and weak inversion to produce valid results. This equation has, in fact, been used to model white noise in all regions of inversion¹⁵⁶ with Q'_I as obtained from the charge sheet model. Interpolation models have also been used for this purpose.^{55,57} An example will be shown in Table 8.1 at the end of this chapter.

INDUCED GATE NOISE. The above results have been based on random fluctuations of the potential in the channel. These fluctuations are coupled to the gate terminal

†It would appear that this equivalence does not hold as V_{DS} is reduced, since the right-hand side of (8.5.27) increases (in agreement with experiment^{139,141}) whereas the expression $2qI_{DS}$ predicts a decreasing value with decreasing V_{DS} (in fact, that value would be 0 at $V_{DS} = 0$, a physically impossible result). This would tend to favor the assumption that the noise in weak inversion is of thermal origin.^{139,141} This discrepancy between the two theories is, however, removed if one views I_{DS} as the superposition of two components, one associated with the drain and one with the source^{148,155} (a similar superposition view is encountered in bipolar transistor theory). Thus, from (4.6.3) we can write $I_{DS} = I_1 - I_2$, where $I_1 = -(W/L)\mu\phi_t Q'_{I0}$ and $I_2 = -(W/L)\mu\phi_t Q'_{IL}$. If each of I_1 and I_2 is assumed to have shot noise (which has zero average value) and the two noise components are assumed uncorrelated, their mean square values will add. Using then equations from Sec. 4.6 it is easy to show (Prob. 8.27) that the combined power spectral density of the shot noise in the two currents is given *identically* by the right-hand side of (8.5.28) for *all* values of V_{DS} . In this way, both the shot noise assumption and the thermal noise assumption are seen to produce identical results for all bias points in weak inversion. A similar observation for the noise in a zero-biased pn junction is well known.¹⁰⁴

through the oxide capacitance, and they “induce” a minute noise current in the gate terminal even if all external voltages are fixed.^{108,111,103,104} This phenomenon is usually neglected in models intended for operation at low or medium frequencies (which are the subject of this chapter). Induced gate noise is considered in Chap. 9.

8.5.3 Flicker Noise

As shown in Fig. 8.33, at low frequencies flicker noise is dominant. This noise is also called “ $1/f$ ” noise, because the power spectral density is nearly proportional to the inverse of the frequency. Flicker noise in MOS transistors has been the subject of intensive studies for several decades. There are several theories for the origin of this noise, with involved physics and sometimes conflicting conclusions, and several remaining unresolved issues. Here we will only summarize the main conclusions of the two dominant theories; the reader is referred to the literature for more information.^{106,160–222}

The first theory attributes the origin of flicker noise to the random fluctuation of the number of carriers in the channel, due to fluctuations in the surface potential; the latter are in turn caused by trapping and releasing of carriers by traps located near the Si-SiO₂ interface.^{160,163,170,171,173,175,194,213} The characteristic times involved in this process cover a very wide range, and when very large numbers of such events are considered, it can be shown that a power spectral density nearly proportional to the inverse of the frequency results.^{103,104} The detailed physics is complicated; here we will give some intuitive arguments to help make plausible the dependence of flicker noise on device parameters. The effects discussed above can be thought of as randomly varying the effective interface charge density Q'_o , thus modulating the flatband voltage through the term Q'_o/C'_{ox} in (2.2.6).²⁰⁷ This is equivalent to a noise voltage in series with the gate, and proportional to $1/C'_{ox}$; the mean square value of this noise is proportional to $(1/C'_{ox})^2$. Flicker noise results from the superposition of such variations due to many traps; the larger the gate area WL , the more the effects of these variations tend to average out, and the smaller the resulting flicker noise. Using detailed physical considerations, the power spectral density of the equivalent noise voltage appearing in series with the gate can be shown to be^{194,207}

$$S_{vf}(f) = \frac{K_1}{C_{ox}^{\prime 2}} \frac{1}{WL} \frac{1}{f^c} \quad (8.5.29)$$

where, for n -channel devices, the exponent c is between 0.7 and 1.2, and K_1 is a quantity independent of bias (see below) but dependent on fabrication details; it is generally found that cleaner fabrication processing can result in lower values for K_1 . Since the equivalent noise voltage appears in series with the gate, one can determine its ef-

fect on the drain current by multiplying by g_m ; thus, instead of an equivalent noise voltage, we can consider an equivalent drain noise current with power spectra density

$$S_{if}(f) = g_m^2 S_{vf}(f) \quad (8.5.30)$$

This represents the contribution of flicker noise to the total power spectral density of the noise current, as shown in (8.5.6).

A second theory attributes flicker noise to mobility fluctuations, due to carrier interactions with lattice fluctuations.^{189,190,191,219,213} Results based on this theory suggest that the power spectral density for the equivalent input noise voltage is

$$S_{vf}(f) = \frac{K(V_{GS})}{C'_{ox}} \frac{1}{WL} \frac{1}{f} \quad (8.5.31)$$

where $K(V_{GS})$ is a bias-dependent quantity (see below). The inverse proportionality with C'_{ox} is not universally accepted. The corresponding noise current power spectral density can again be found from (8.5.30).

The above results are generally used for all regions of inversion. Some research suggests that flicker noise is due to both carrier number fluctuations and mobility fluctuations.^{193,208} After all, the carrier number fluctuation theory talks about a randomly varying charge at traps near the interface, and we have seen that charge can affect mobility through "Coulomb scattering" (Sec. 4.10); so it is not surprising that both effects may be present and correlated in a given device.²⁰⁴ However, one or the other effect may be dominant. For example, flicker noise in n -channel devices is today often attributed to carrier number fluctuations; such devices may follow (8.5.29), with typical K_1 values between 5×10^{-31} and 1×10^{-30} C²·cm⁻². On the other hand, the flicker noise of p -channel devices is often attributed to mobility fluctuations and is predicted by (8.5.31), with typical values of $K(V_{GS})$ of the order of 6×10^{-26} to 2×10^{-23} V²·F at $|V_{GS} - V_T|$ of about 1 V. For p -channel devices it is found^{189,216} that, in strong inversion, $K(V_{GS})$ increases approximately linearly with $|V_{GS} - V_T|$ to various degrees, from negligible to rather high [e.g., $K(V_{GS})$ can increase by several hundred percent as V_{GS} is increased over several volts]; however, it has been reported that, in weak inversion, $K(V_{GS})$ increases as $|V_{GS}|$ is decreased.²¹⁶

Flicker noise power spectral densities in p -channel devices are generally found to be significantly less than those of n -channel devices of the same dimensions and fabricated with the same CMOS process (by 1 order of magnitude or more), unless $|V_{GS} - V_T|$ is very large, in which case the two types of devices may give similar noise.²¹⁶ For buried-channel pMOS devices (Sec. 5.4.2), the low value of flicker noise observed is attributed to the fact that the channel is farther away from the

Si-SiO₂ interface, and is thus less affected by interface traps. The power spectral density of some pMOS buried-channel devices has been reported to be up to 2 orders of magnitude lower than those of pMOS surface-channel devices,^{198,219} but this is not always observed.

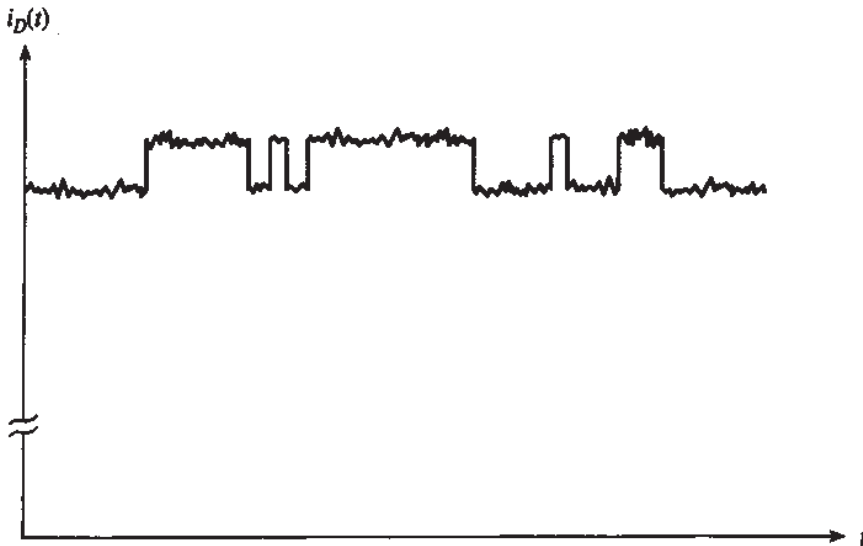
8.5.4 Small-Dimension Effects

The results presented so far are for long-channel devices. Short-channel devices can be significantly noisier than the above formulas would predict.^{134,138,223–233} One reason these formulas fail in such cases is that they do not include short-channel effects such as velocity saturation. Another is that the high electric fields present in such devices near the drain produce “hot carriers” in that vicinity, for sufficiently large V_{DS} values (Sec. 6.6). The effect of these can be modeled by an equivalent “carrier temperature” that is higher than the lattice temperature,¹⁰³ so it is not surprising that the thermal noise of short-channel devices can be high. A “noise excess factor” can be defined as the factor by which the noise current power spectral density is higher than that predicted by long-channel theory. Both measurements^{147,224} and models²³¹ show that noise excess factors of 2 to 5, or even more, can be expected.

As we have seen, hot carriers are also responsible for the drain-to-substrate current, I_{DB} . This current shows noise of its own;²²⁴ at low substrate current values, this has been found to be shot noise, with power spectral density of $2qI_{DB}$. At higher I_{DB} values, excess noise is found in this current, too. Since I_{DB} produces a voltage drop across the substrate resistance, its noise modulates the effective substrate potential; the fluctuations in the latter are coupled to the drain current through the substrate transconductance g_{mb} and can contribute to an increase in the observed drain current noise.

Hot carriers can also compromise the quality of the oxide near the interface, and introduce additional traps near the interface in the area close to the drain. Consider a device in which such traps have been generated, due to, say, operation at or near punchthrough for several hours. When the drain voltage is reduced and the device is operated in the normal saturation region, this does not result in a noise increase since these traps are located in the pinchoff region, which does not affect significantly the value of the current. If, however, the device is operated in nonsaturation, that region of the channel affects the channel charge and the current, and the extra traps cause an increase in flicker noise.^{223,225–229,232} The same happens if the above device is operated in reverse saturation (with the source and drain terminals interchanged), in which case the damaged area will lie next to the new “source.” Excess flicker noise of over an order of magnitude can be observed in this case.

Consider now a device with negligible hot carrier effects. If the gate area WL is very small (a fraction of $1\ \mu\text{m}^2$), the flicker noise can be expected to be high, as can be seen from the formulas in the previous section. This is because there will now be only a few traps which can exchange charge with the channel, and their individual effects will be noticed, rather than tending to average out as in the case of large gate area. It is, in fact, possible, that only a single trap of this type exists in a very small device. Then, as it captures and releases charge, abrupt changes in the drain current

**FIGURE 8.38**

Typical variation of the drain current in a device with extremely small gate area, due to random telegraph noise.

can be noticed, as shown in Fig. 8.38; these changes will be on top of the more common noise variation.^{234–241} This is reminiscent of the so-called “random telegraph signals” (RTS) in communication theory, and for this reason is often referred to as random telegraph noise. The variation of the drain current due to this effect can be significant (e.g., 0.1 percent, or more, of the dc current value). It is interesting that the flicker noise observed in large-dimension devices has been viewed as a superposition of many RTS waveforms, of the type shown in Fig. 8.38.²³⁵

Noise in small-dimension devices is the subject of current research. Although several models have been proposed, they have not been widely tested yet, and their incorporation in circuit simulators is far from complete.

8.5.5 Equivalent-Circuit Model

Adding a noise current source to the small-signal equivalent circuit of Fig. 8.17 produces the model in Fig. 8.39. The power spectral density of the noise source is given by (8.5.6). When using such models to calculate noise in circuits, one has to work with mean square or root mean square values of noise currents and voltages, rather than with the noise currents and voltages themselves, and this requires some care. Convenient techniques for such calculations are given in several texts.^{103–105}

The extrinsic resistances discussed in Sec. 8.4 exhibit thermal noise. One can model this effect as discussed in Sec. 8.5.1. For certain device structures, the polysilicon gate resistance can be a significant contributor to noise.^{145,159} The same is true for the substrate resistance,^{146,147} the noise of which is coupled to the drain current through the substrate transconductance g_{mb} . One way to reduce this effect is to decrease g_{mb} which, as follows from Sec. 8.2, can be achieved by increasing the source-substrate bias V_{SB} .¹⁴⁷

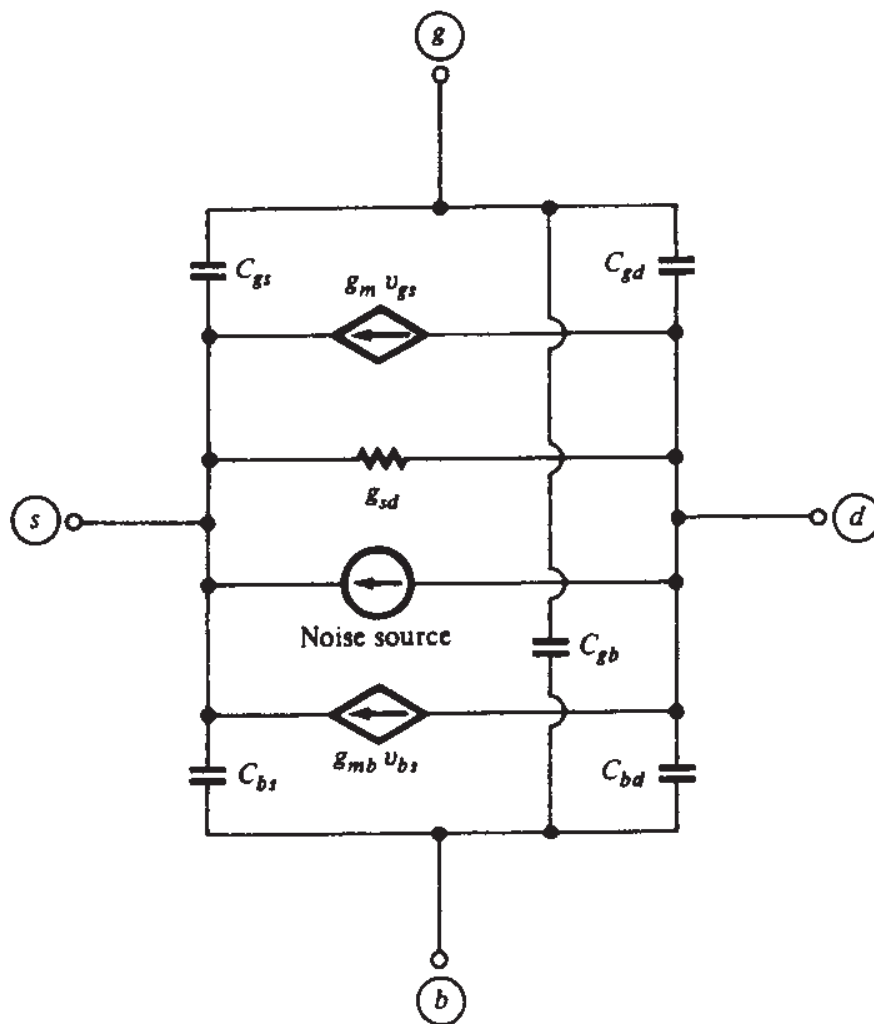


FIGURE 8.39

A small-signal equivalent circuit for the intrinsic part of a transistor, with a noise current source added; the power spectral density of this source is given by (8.5.6).

8.6 GENERAL MODELS

General expressions for small-signal parameters can be developed for the simplified charge-sheet model⁵¹⁻⁵³ discussed in Sec. 4.3.2, as well as for the EKV model^{54-57a} discussed in Sec. 4.8. An example of a set of expressions valid in weak, moderate, and strong inversion *saturation* regions is shown in Table 8.1. Except for channel length modulation, short-channel effects are not included. All symbols not defined in the table are as used in the corresponding sections of this chapter.

The value of n shown is a compromise for medium accuracy in all regions, and it is adequate for simple estimates. In weak inversion, better accuracy is obtained if n is modeled as in (3.4.25). In strong inversion, if the expected voltage ranges are significant, better accuracy for I_{DS} is obtained if n is replaced by α as used in Sec. 4.5.3. However, unless the transition between n and α is made gradually as the moderate inversion is crossed, discontinuities will occur.

TABLE 8.1

A saturation region long-channel nMOS model valid in all regions of inversion†

Drain current:
$$I_{DS} = I_Z \ln^2 \left[1 + \exp \left(\frac{V_{GS} - V_T}{2n\phi_t} \right) \right]$$

Small-signal conductances:

$$g_m = \frac{I_{DS}}{n\phi_t} \frac{1}{f(u)}$$

$$g_{mb} = (n - 1)g_m$$

$$g_{sd} = \frac{I_{DS}}{V_A}$$

Small-signal *intrinsic* capacitances:

$$C_{gs} = WLC'_{ox} \left[\frac{3}{2} + \frac{f(u)}{u} \right]^{-1}$$

$$C_{bs} = (n - 1)C_{gs}$$

$$C_{gb} = WLC'_{ox} \frac{n - 1}{n} \left[1 - \left(\frac{3}{2} + \frac{f(u)}{u} \right)^{-1} \right]$$

$$C_{gd} = C_{bd} = 0$$

Noise power spectral densities:

$$S_i(f) = 4kT \left\{ \frac{1}{2} + \frac{1}{6} \frac{u}{[f(u)]^2} \right\} n g_m$$

$$S_{vf}(f) = \frac{K_1}{C'^2_{ox}} \cdot \frac{1}{WL} \cdot \frac{1}{f}$$

$$I_Z = \frac{W}{L} \mu C'_{ox} (2n\phi_t^2)$$

$$V_T = V_{T0} + \gamma \left(\sqrt{V_{SB} + \phi_0} - \sqrt{\phi_0} \right)$$

$$n = 1 + \frac{\gamma}{2\sqrt{V_{SB} + \phi_0}}$$

$$u = \frac{I_{DS}}{I_Z}$$

$$f(u) = \frac{1}{2} \left(\sqrt{1 + 4u} + 1 \right)$$

†The current equation²⁴² has been adopted from Refs. 54–57a for using the source as a reference and for making possible the inclusion of an accurately evaluated V_T . The small-signal and noise expressions are from Refs. 55 and 57. The expression for $f(u)$ is from Refs. 52 and 53.

REFERENCES

1. H. J. K. Ihantola and J. L. Moll, "Design theory of a surface field-effect transistor," *Solid-State Electronics*, vol. 7, pp. 423-430, June 1964.
2. C. T. Sah, "Characteristics of the metal-oxide semiconductor transistors," *IEEE Transactions on Electron Devices*, vol. ED-11, pp. 324-345, July 1964.
3. S. R. Hofstein and G. Warfield, "Carrier mobility and current saturation in the MOS transistor," *IEEE Transactions on Electron Devices*, vol. ED-12, pp. 129-138, March 1965.
4. V. K. G. Reddi and C. T. Sah, "Source to drain resistance beyond pinch-off in metal-oxide semiconductor transistors (MOST)," *IEEE Transactions on Electron Devices*, vol. ED-12, pp. 139-141, March 1965.
5. M. H. White and R. C. Gallagher, "Metal oxide semiconductor (MOS) small-signal equivalent circuits," *Proceedings of the IEEE*, vol. 53, pp. 314-315, 1965.
6. C. T. Sah and H. C. Pao, "The effects of fixed bulk charge on the characteristics of metal-oxide semiconductor transistors," *IEEE Transactions on Electron Devices*, vol. ED-13, pp. 393-409, April 1966.
7. M. B. Das, "Charge-control analysis of M.O.S. and junction-gate field-effect transistors," *IEE Proceedings*, vol. 113, pp. 1565-1570, October 1966.
8. F. A. Lindholm, R. J. Balda, and J. L. Clements, "Characterization of the four-terminal MOS transistor for digital and linear applications," *Digest of Technical Papers*, International Electronics Conference, Toronto, pp. 116-117, 1967.
9. M. B. Das, "Dependence of the characteristics of MOS transistors on the substrate resistivity," *Solid-State Electronics*, vol. 11, pp. 305-322, March 1968.
10. J. E. Schroeder and R. S. Muller, "IGFET analysis through numerical solution of Poisson's equation," *IEEE Transactions on Electron Devices*, vol. ED-15, no. 12, pp. 954-961, December 1968.
11. T. L. Chiu and C. T. Sah, "Correlation experiments with a two-section model theory of the saturation drain conductance of MOS transistors," *Solid-State Electronics*, vol. 11, pp. 1149-1163, 1968.
12. D. Frohman-Bentchkowsky and A. S. Grove, "Conductance of MOS transistors in saturation," *IEEE Transactions on Electron Devices*, vol. ED-16, pp. 108-113, January 1969.
13. R. S. C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley-Interscience, New York, 1970.
14. G. Baum and H. Beneking, "Drift velocity saturation in MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-17, pp. 481-482, 1970.
15. D. J. Hamilton, F. A. Lindholm, and A. H. Marshak, *Principles and Applications of Semiconductor Device Modeling*, Holt, Rinehart, and Winston, New York, 1971.
16. J. E. Meyer, "MOS models and circuit simulation," *RCA Review*, vol. 32, pp. 42-63, March 1971.
17. G. Merckel, J. Borel, and N. Z. Cupcea, "An accurate large-signal MOS transistor model for use in computer-aided design," *IEEE Transactions on Electron Devices*, vol. ED-19, pp. 681-690, May 1972.
18. A. Popa, "An injection level dependent theory of the MOS transistor in saturation," *IEEE Transactions on Electron Devices*, vol. ED-19, pp. 774-781, June 1972.
19. E. Vittoz and J. Fellrath, "MOS analog integrated circuits based on weak inversion operation," *IEEE Journal of Solid-State Circuits*, vol. SC-12, pp. 224-231, June 1977.
20. J. Fellrath and E. Vittoz, "Small signal model of MOS transistors in weak inversion," *Proc. Journées d'Electronique 1977*, Session C4, Ecole Polytechnique Fédérale de Lausanne, Switzerland, pp. 315-324, 1977.
21. P. Rossel, H. Martinot, and G. Vassilief, "Accurate two-sections model for MOS transistors in saturation," *Solid-State Electronics*, vol. 19, pp. 51-56, January 1976.
22. F. M. Klaassen, "A MOS model for computer-aided design," *Philips Research Reports*, vol. 31, pp. 71-83, 1976.
23. Y. A. El-Mansy and A. R. Boothroyd, "A simple two-dimensional model for IGFET operation in the saturation region," *IEEE Transactions on Electron Devices*, vol. ED-24, pp. 254-262, March 1977.

24. G. W. Taylor, "The effects of two-dimensional charge sharing on the above-threshold characteristics of short-channel devices," *Solid-State Electronics*, vol. 22, pp. 701–717, 1979.
25. H. C. Poon, " V_{th} and beyond," presented at the Workshop on Device Modelling for VLSI, Burlingame, California, March 29, 1979; also L. Cong, Bell Laboratories, private communication.
26. T. Poorter and J. H. Satter, "A D.C. Model for an MOS-transistor in the saturation region," *Solid-State Electronics*, vol. 23, pp. 765–772, 1980.
27. G. Cardinali, S. Graffi, M. Impronta, and G. Masetti, "DC MOSFET model for analogue circuit simulation employing process-empirical parameters," *IEE Proceedings*, vol. 129, part I, pp. 61–66, April 1982.
28. S. Liu and L. W. Nagel, "Small-signal MOSFET models for analog circuit design," *IEEE Journal of Solid-State Circuits*, vol. SC-17, pp. 983–998, December 1982.
29. C. Turchetti, G. Masetti, and Y. Tsividis, "On the small-signal behavior of the MOS transistor in quasi-static operation," *Solid-State Electronics*, vol. 26, pp. 941–949, 1983.
30. M. E. Nokali and H. Miranda, "A simple model for the MOS transistors in saturation," *Solid-State Electronics*, vol. 29, pp. 591–596, 1986.
31. C. N.-Duc, S. Cristoloveanu, and G. Ghibaudo, "A three-piece model of channel length modulation in submicrometer MOSFETs," *Solid-State Electronics*, vol. 31, pp. 1057–1063, 1988.
32. F. S. Shoucair, "A semi-empirical model of the MOSFET's small-signal drain conductance in saturation for analog circuit design," *IEEE Transactions on Electron Devices*, vol. 39, pp. 1246–1248, May 1992.
33. J. H. Huang, Z. H. Liu, M. C. Jeng, P. K. Ko, and C. Hu, "A physical model for MOSFET output resistance," *Digest, International Electron Devices Meeting*, pp. 569–579, 1992.
34. M. Fujishima and K. Asada, "A nonpinchoff gradual channel model for deep-submicron MOSFET's," *IEEE Transactions on Electron Devices*, vol. 40, pp. 1883–1885, October 1993.
35. R. M. D. A. Velghe, D. B. M. Klaassen, and F. M. Klaassen, "Compact MOS modeling for analog circuit simulation," *Digest, International Electron Devices Meeting*, pp. 485–488, December 1993.
36. M. Miura-Mattausch, "Analytical MOSFET model for quarter micron technologies," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 13, pp. 610–615, May 1994.
37. J. K. Seon and K. S. Yoon, "A precision output conductance model for analog CMOS circuit simulations," *Digest, International Electron Devices Meeting*, pp. 1584–1587, December 1995.
38. W. R. Bandy and R. S. Winton, "A new approach for modeling the MOSFET using a simple, continuous analytical expression for drain conductance which includes velocity-saturation in a fundamental way," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, pp. 475–483, May 1996.
39. Y. Cheng, M.-C. Jeng, Z. Liu, K. Chen, M. Chan, C. Hu, and P. K. Ko, "An investigation on the robustness, accuracy and simulation performance of a physics-based deep-submicrometer BSIM model for analog/digital circuit simulation," *IEEE Custom Integrated Circuits Conference*, pp. 321–324, 1996.
40. Y. G. Chen and J. B. Kuo, "A unified triode/saturation model with an improved continuity in the output conductance suitable for CAD of VLSI circuits using deep sub-0.1 μm NMOS devices," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, pp. 256–258, February 1996.
41. Y. Cheng, M.-C. Jeng, Z. Liu, J. Huang, M. Chan, K. Chen, P. K. Ko, and C. Hu, "A physical and scalable I - V model in BSIM3v3 for analog/digital circuit simulation," *IEEE Transactions on Electron Devices*, vol. 44, pp. 277–287, February 1997.
42. H. C. de Graaff and F. M. Klaassen, *Compact Transistor Modelling for Circuit Design*, Springer-Verlag, Vienna, 1990.
43. N. Arora, *MOSFET Models for VLSI Circuit Simulation—Theory and Practice*, Springer-Verlag, Vienna, 1993.
44. K. Lee, M. Shur, T. A. Fjeldly, and T. Ytterdal, *Semiconductor Device Modeling for VLSI*, Prentice-Hall, Englewood Cliffs, N.J., 1993.

45. Y. Tsividis and G. Masetti, "Problems in precision modeling of the MOS transistor for analog applications," *IEEE Transactions on Computer-Aided Design*, vol. CAD-3, pp. 72–79, January 1983.
46. C. C. McAndrew, B. K. Bhattacharyya, and O. Wing, "A single-piece C_{∞} -continuous MOSFET model including subthreshold conduction," *IEEE Electron Device Letters*, vol. 12, pp. 565–567, October 1991.
47. S. Cserveny, "MOS small-signal conductances in the weak avalanche multiplication region," *IEEE Transactions on Electron Devices*, vol. 39, pp. 1233–1235, May 1992.
48. Y. Tsividis, "Moderate inversion in MOS devices," *Solid-State Electronics*, vol. 25, pp. 1099–1104, 1982; see also Erratum, *ibid.*, vol. 26, p. 823, 1983.
49. H.-S. Wong, M. H. White, T. J. Krutsick, and R. V. Booth, "Modeling of transconductance degradation and extraction of threshold voltage in thin oxide MOSFET's," *Solid-State Electronics*, vol. 30, pp. 953–968, 1987.
50. J. R. Brews, "A charge sheet model of the MOSFET," *Solid-State Electronics*, vol. 21, pp. 345–355, 1978.
51. A. I. A. Cunha, M. C. Schneider, and C. G.-Montor, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electronics*, vol. 38, pp. 1945–1952, November 1995.
52. A. I. A. Cunha, *Um Modelo do Transistor MOS para Projecto de Cicuitos Integrados*, Ph.D. thesis, Universidade Federal de Santa Catarina, December 1996 (in Portuguese).
53. A. I. A. Cunha, O. C. Gouveia-Filho, M. C. Schneider, and C. Galup-Montoro, "A current-based model for the MOS transistor," *Proceedings 1997 International Symposium on Circuits and Systems*, pp. 1608–1611, Hong Kong, June 1997.
54. H. J. Oguey and S. Cserveny, "MOS Modeling at Low Current Density," Summer Course on Process and Device Modeling, K. U. Leuven, The Netherlands, 1983.
55. C. Enz, *High Precision CMOS Micropower Amplifiers*, Ph.D. thesis no. 802, Ecole Polytechnique Fédéral de Lausanne, Switzerland, 1989.
56. E. A. Vittoz, "Micropower Techniques," in J. E. Franca and Y. Tsividis (editors), *Design of Analog-Digital VLSI Circuits for Telecommunications and Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1994.
57. C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog Integrated Circuits and Signal Processing*, vol. 8, pp. 83–114, July 1995.
- 57a. M. Bucher, C. Lallement, C. Enz, F. Theodoloz, and F. Krummenacher, "Scalable GM/I based MOSFET model," Proc. International Semiconductor Device Research Symposium, Charlottesville, VA, pp. 615–618, December 1997.
58. Y. P. Tsividis, "Reaction between incremental intrinsic capacitances and transconductances in MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 946–948, May 1980.
59. D. E. Ward, "Charge-based modeling of capacitance in MOS transistors," *Technical Report G201-11*, Integrated Circuits Laboratory, Stanford University, California, June 1981.
60. J. J. Paulos, D. A. Antoniadis, and Y. P. Tsividis, "Measurement of intrinsic capacitances of MOS transistors," *Technical Digest*, IEEE International Solid-State Circuits Conference, San Francisco, pp. 238–239, February 1982.
61. J. J. Paulos and D. A. Antoniadis, "Limitations of quasi-static capacitance models for the MOS transistor," *IEEE Electron Device Letters*, vol. EDL-4, pp. 221–224, July 1983.
62. C. Turchetti, G. Masetti, and Y. Tsividis, "On the small-signal behavior of the MOS transistor in quasi-static operation," *Solid-State Electronics*, vol. 26, pp. 941–949, 1983.
63. G. I. Serhan and S.-Y. Yu, "A simple charge-based model for MOS transistor capacitances: A new production tool," *IEEE Transactions on Computer-Aided Design*, vol. CAD-2, p. 48, January 1983.
64. E. W. Greeneich, "An analytical model for the gate capacitance of small-geometry MOS structures," *IEEE Transactions on Electron Devices*, ED-30, pp. 1838–1839, 1983.
65. J. Oristian, H. Iwai, J. Walker, and R. Dutton, "Small geometry MOS transistor capacitance measurement method using simple on-chip circuits," *IEEE Electron Device Letters*, vol. EDL-5, pp. 395–397, October 1984.

66. B. J. Sheu and P. K. Ko, "An analytical model for intrinsic capacitances of short-channel MOS-FETs," *Technical Digest, International Electron Devices Meeting*, pp. 300–303, San Francisco, 1984.
67. H. Iwai, J. E. Oristian, J. T. Walker, and R. W. Dutton, "A scalable technique for the measurement of intrinsic MOS capacitance with atto-Farad resolution," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 344–356, February 1985.
68. J. J. Paulos and D. A. Antoniadis, "Measurement of minimum-geometry MOS transistor capacitances," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 357–363, February 1985.
69. H. Iwai, M. R. Pinto, C. S. Rafferty, J. E. Oristian, and R. W. Dutton, "Velocity saturation effect on short-channel MOS transistor capacitance," *IEEE Electron Device Letters*, vol. EDL-6, pp. 120–122, March 1985.
70. T. Shima, "Table lookup MOSFET capacitance model for short-channel devices," *IEEE Transactions on Computer-Aided Design*, vol. CAD-5, pp. 624–632, October 1986.
71. H. Iwai, M. R. Pinto, C. S. Rafferty, J. E. Oristian, and R. W. Dutton, "Analysis of velocity saturation and other effects on short-channel MOS transistor capacitances," *IEEE Transactions on Computer-Aided Design*, vol. CAD-6, pp. 173–184, March 1987.
72. Y. Ohkura, T. Toyabe, and H. Masuda, "Analysis of MOSFET capacitances and their behavior at short-channel lengths using an AC device simulator," *IEEE Transactions on Computer-Aided Design*, vol. CAD-6, pp. 423–430, March 1987.
73. F. Riedel, "Kleinsignalmodell für kurz- und schmalkanal-MOS-transistoren," *AEÜ*, vol. 41, pp. 13–20, 1987.
74. Y.-T. Yeow, "Measurement and numerical modeling of short-channel MOSFET gate capacitances," *IEEE Transactions on Electron Devices*, vol. ED-35, pp. 2510–2520, December 1987.
75. H. Masuda, Y. Aoki, J. Mano, and O. Yamashiro, "MOSTSM: a physically based charge conservative MOSFET model," *IEEE Transactions on Computer-Aided Design*, vol. 7, pp. 1229–1236, December 1988.
76. B.J. Sheu and P.-K. Ko, "Measurement and modeling of short-channel MOS transistor gate capacitances," *IEEE Journal of Solid-State Circuits*, vol. SC-22, pp. 464–472, June 1987.
77. B.J. Sheu, W.-J. Hsu, and P. K. Ko, "An MOS transistor charge model for VLSI design," *IEEE Transactions on Computer-Aided Design*, vol. 7, pp. 520–527, April 1988.
78. R. Gharabagi and M. E. Nokali, "A model for the intrinsic gate capacitances of short channel MOS-FETs," *Solid-State Electronics*, vol. 32, pp. 57–63, 1989.
79. K. A. Sakallah, Y.-T. Yen, and S. S. Greenberg, "A first-order charge conserving MOS capacitance model," *IEEE Transactions on Computer-Aided Design*, vol. 9, pp. 99–108, January 1990.
80. W. Budde and W. H. Lamfried, "A charge-sheet capacitance model based on drain current modeling," *IEEE Transactions on Electron Devices*, vol. 37, pp. 1678–1687, July 1990.
81. R. Gharabagi and M. E. Nokali, "An analytical model for the capacitances in short-channel MOS-FETs," *Solid-State Electronics*, vol. 33, pp. 235–241, 1990.
82. C. S. Oh, W. H. Chang, B. Davari, and Y. Taur, "Voltage dependence of the MOSFET gate-to-source/drain overlap," *Solid-State Electronics*, vol. 33, pp. 1650–1652, 1990.
83. R. Gharabagi, and M. A. E.-Nokali, "A charge-based model for short-channel MOS transistor capacitances," *IEEE Transactions on Electron Devices*, vol. 37, pp. 1064–1073, April 1990.
84. H.-J. Park, P. K. Ko, and C. Hu, "A charge sheet capacitance model of short channel MOSFET's for SPICE," *IEEE Transactions on Computer-Aided Design*, vol. 10, pp. 376–389, March 1991.
85. A. A.-Kushaa and M. E. Nokali, "Modeling subthreshold capacitances of MOS transistors," *Solid-State Electronics*, vol. 35, pp. 45–49, 1992.
86. P. Klein, K. Hoffmann, and B. Lemaitre, "Description of the bias dependent overlap capacitance at LDD MOSFETs for circuit applications," *Digest, International Electron Devices Meeting*, pp. 493–496, December 1993.
87. K.-M. Rho, K. Lee, M. Shur, and T. A. Fjeldly, "Unified quasi-static MOSFET capacitance model," *IEEE Transactions on Electron Devices*, vol. 40, pp. 131–135, January 1993.
88. T. Smedes and F. M. Klaassen, "Influence of channel series resistances on dynamic MOSFET behaviour," *Solid-State Electronics*, vol. 37, pp. 251–254, 1994.

89. J. Katzenelson and A. Unikovski, "A network charge-oriented MOS transistor model," *International Journal of High Speed Electronics and Systems*, vol. 6, pp. 285–316, 1995.
90. D.-H. Cho, S.-M. Kang, K.-H. Kim, and S.-H. Lee, "An accurate intrinsic capacitance modeling for deep submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, vol. 42, pp. 540–548, March 1995.
91. V. I. Kol'dyaev, A. Clerix, R. M. Arteaga, and L. Deferm, "Characterisation of the overlap capacitance of submicron LDD MOSFETs," *Proceedings of the 25th European Solid State Device Research Conference*, pp. 757–760, The Netherlands, September 1995.
92. B. Riccò, R. Versari, and D. Esseni, "A novel method to characterize parasitic capacitances in MOSFET's," *IEEE Electron Device Letters*, vol. 16, pp. 485–487, November 1995.
93. T. Manku, "A methodology for measuring the gate-drain capacitance of CMOS devices," *IEEE Electron Device Letters*, vol. 17, pp. 312–314, June 1996.
94. C. H. Wang, "Identification and measurement of scaling-dependent parasitic capacitance of small-geometry MOSFET's," *IEEE Transactions on Electron Devices*, vol. 43, p. 965, June 1996.
95. R. Shrivastava and K. Fitzpatrick, "A simple model for the overlap capacitance of a VLSI MOS device," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 1870–1875, December 1982.
96. K. S. Krisch, J. D. Bude, and L. Manchanda, "Gate capacitance attenuation in MOS devices with thin gate dielectrics," *IEEE Electron Device Letters*, vol. 17, pp. 521–524, November 1996.
97. P. Klein, "A compact charge LDD MOSFET model," *IEEE Transactions on Electron Devices*, vol. 44, pp. 1483–1490, September 1997.
98. H. Murtman and D. Widmann, "Current crowding on metal contacts to planar devices," *IEEE Transactions Electron Devices*, vol. ED-16, pp. 1022–1026, 1969.
99. H. H. Berger, "Models for contacts to planar devices," *Solid-State Electronics*, pp. 145–158, 1972.
100. G. Baccarani and G. A. Sai-Halasz, "Spreading resistance in submicron MOSFETs," *IEEE Electron Device Letters*, vol. EDL-4, pp. 27–29, 1983.
101. K. K. Ng, R. J. Bayruns, and S. C. Fang, "The spreading resistance of MOSFETs," *IEEE Electron Device Letters*, vol. EDL-6, pp. 195–197, 1985.
102. K. K. Ng and W. T. Lynch, "Analysis of the Series Resistance of MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 965–972, 1986.
103. A. van der Ziel, *Noise in Solid State Devices and Circuits*, Wiley-Interscience, New York, 1986.
104. A. Ambrozy, *Electronic Noise*, McGraw-Hill, New York, 1982.
105. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1993.
- 105a. C. A. Desoer and E. S. Kuh, *Basic Circuit Theory*, McGraw-Hill, New York, 1969.
106. C. T. Sah, "Theory and experiments on the $1/f$ surface noise of MOS insulated-gate field-effect transistors," *IEEE Transactions on Electron Devices (Abstract)*, vol. ED-11, p. 534, November 1964.
107. A. G. Jordan and N. A. Jordan, "Theory of noise in metal oxide semiconductor devices," *IEEE Transactions on Electron Devices*, vol. ED-12, pp. 148–156, March 1965.
108. H. Johnson, "Noise in field-effect transistors," chapter 6 in *Field-Effect Transistors*, T. Wallmark and H. Johnson (editors), Prentice-Hall, Englewood Cliffs, N.J., 1966.
109. S. M. Bozic, "Noise in the metal oxide semiconductor transistor," *Electronic Engineering*, vol. 38, pp. 40–41, 1966.
110. C. T. Sah, S. Y. Wu, and F. H. Hielscher, "The effects of fixed bulk charge on the thermal noise in metal-oxide-semiconductor transistors," *IEEE Transactions on Electron Devices*, vol. ED-13, pp. 410–414, April 1966.
111. M. Shoji, "Analysis of high-frequency thermal noise of enhancement mode M.O.S. field-effect transistors," *IEEE Transactions on Electron Devices*, vol. ED-13, pp. 520–524, June 1966.
112. I. Flinn, G. Bew, and F. Berz, "Low frequency noise in M.O.S. field effect transistors," *Solid-State Electronics*, vol. 10, pp. 833–845, August 1967.
113. J. Mavor, "Noise parameters for metal-oxide-semiconductor transistors," *IEE Proceedings*, vol. 113, pp. 1463–1467, September 1967.
114. F. M. Klaassen and J. Prins, "Thermal noise of M.O.S. transistors," *Philips Research Reports*, vol. 22, pp. 505–514, October 1967.

115. R. Paul, "Thermisches Rauschen von MOS-Transistoren," *Nachrichtentechnik*, vol. 17, pp. 458–466, December 1967.
116. S. Y. Wu, "Theory of the generation-recombination noise in MOS transistors," *Solid-State Electronics*, vol. 11, pp. 25–32, 1968.
117. A. Leupp and M. J. O. Strutt, "Noise behavior of the MOSFET at VHF and UHF," *Electronics Letters*, vol. 4, pp. 313–314, July 1968.
118. H. E. Halladay and A. van der Ziel, "Test of the thermal noise hypothesis in MOSFETS," *Electronics Letters*, vol. 4, pp. 366–367, August 23, 1968.
119. L. D. Yau and C. T. Sah, "Theory and experiments of low-frequency generation recombination noise in MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-16, pp. 170–177, February 1969.
120. F. M. Klaassen and J. Prins, "Noise of field-effect transistors at very high frequencies," *IEEE Transactions on Electron Devices*, vol. ED-16, pp. 952–957, 1969.
121. M. Nakahara, "Anomalous low-frequency noise enhancement in silicon MOS transistors," *Proceedings of the IEEE*, vol. 57, pp. 2177–2178, 1969.
122. L. D. Yau and C. T. Sah, "On the excess 'white noise' in MOS transistors," *Solid-State Electronics*, vol. 12, pp. 927–936, 1969.
123. A. van der Ziel, *Noise: Sources, Characterization, and Measurement*, Prentice-Hall, Englewood Cliffs, N.J., 1970.
124. J. W. Haslett and F. N. Trofimenkoff, "Gate noise in MOSFET's at moderately high frequencies," *Solid-State Electronics*, vol. 14, pp. 239–245, 1971.
125. A. van der Ziel, "Noise resistance of FETs in the hot electron regime," *Solid-State Electronics*, vol. 14, pp. 347–350, 1971.
126. P. S. Rao and A. van der Ziel, "Noise and y parameters in MOSFETs," *Solid-State Electronics*, vol. 14, pp. 939–944, 1971.
127. E. W. Kirk, "Induced gate noise in MOSFETs," *Solid-State Electronics*, vol. 14, pp. 945–948, 1971.
128. M. B. Das and J. M. Moore, "Measurements and interpretation of low frequency noise in FETs," *IEEE Transactions on Electron Devices*, vol. ED-21, pp. 247–257, 1974.
129. S. T. Hsu and A. van der Ziel, "Thermal noise in ion-implanted MOSFETs," *Solid-State Electronics*, vol. 18, pp. 509–510, 1975.
130. N. Nakamura, O. Kudoh, and M. Kamoshida, "Noise characteristics of ion-implanted MOS transistors," *Journal of Applied Physics*, vol. 46, pp. 3189–3193, 1975.
131. W. Fichtner, E. Hochmair, and D. Kranzer, "Noise measurements on SOSMOS transistors," *European Solid-State Device Research Conference*, Munich, September, 1976.
132. W. Fichtner and E. Hochmair, "Current-kink noise or n-channel enhancement ESFI MOS SOS transistors," *Electronics Letters*, vol. 13, pp. 675–676, 1977.
133. P. Gentil and S. Chausse, "Low-frequency measurement on silicon-on-sapphire (SOS) MOS transistors," *Solid-State Electronics*, vol. 20, pp. 935–940, 1977.
134. K. Takagi and K. Matsumoto, "Noise in silicon and FET's at high electric fields," *Solid-State Electronics*, vol. 20, pp. 1–3, 1977.
135. W. A. Baril, "High-frequency thermal noise in MOSFETs," *Solid-State Electronics*, vol. 21, pp. 589–592, 1978.
136. J. Fellrath, "Shot noise behavior of subthreshold MOS transistors," *Revue de Physique Appliquée*, vol. 13, pp. 719–723, December 1978.
137. K. Takagi and A. van der Ziel, "Drain noise in MOSFETS at zero drain bias as a function of temperature," *Solid-State Electronics*, vol. 22, pp. 87–88, 1979.
138. K. Takagi and A. van der Ziel, "Excess high frequency noise and flicker noise in MOSFETs," *Solid-State Electronics*, vol. 22, pp. 289–292, 1979.
139. S. T. Liu and A. van der Ziel, "High-frequency noise in weakly inverted metal-oxide-semiconductor transistors," *Applied Physics Letters*, vol. 37, pp. 950–951, 1980.
140. R. P. Jindal and A. van der Ziel, "Effect of transverse electric field on Nyquist noise," *Solid-State Electronics*, vol. 24, pp. 905–906, 1981.
141. G. Reimbold and P. Gentil, "White noise of MOS transistors operating in weak inversion," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 1722–1725, November 1982.

142. H. S. Park and A. van der Ziel, "Noise measurements in ion implanted MOSFETs," *Solid-State Electronics*, vol. 26, pp. 747–751, 1983.
143. A. van der Ziel, R. J. J. Zijstra, H. S. Park, and S. T. Liu, "Alternate explanation of $1/f$ noise in ion-implanted MOSFETs," *Solid-State Electronics*, vol. 26, pp. 927–928, 1983.
144. S. A. Hayat and B. K. Jones, "Thermal noise in inversion layers," *Solid-State Electronics*, vol. 27, pp. 687–688, 1984.
145. R. P. Jindal, "Noise associated with distributed resistance of MOSFET gate structures in integrated circuits," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 1505–1509, October 1984.
146. R. P. Jindal, "Distributed substrate resistance noise in fine-line NMOS field-effect transistors," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 2450–2453, November 1985.
147. R. P. Jindal, "High frequency noise in fine line NMOS field effect transistors," *Technical Digest, IEEE International Electron Devices Meeting*, Washington, D.C., pp. 68–71, 1985.
148. E. A. Vittoz, private communication, 1985.
149. S. A. Hayat and B. K. Jone, "Thermal noise in inversion layers," *Solid-State Electronics*, vol. 27, pp. 687–688, 1984.
150. E. N. Wu and A. van der Ziel, "On the influence of substrate doping on the input conductance and the induced gate noise in MOSFETS," *Solid-State Electronics*, vol. 27, pp. 945–946, 1984.
151. A. A. Abidi, "High-frequency noise measurements on FET's with small dimensions," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 1801–1805, November 1986.
152. K.K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "A physics-based MOSFET noise model for circuit simulators," *IEEE Transactions on Electron Devices*, vol. 37, pp. 1323–1333, May 1990.
153. L.-J. Pu and Y. Tsvividis, "Small-signal parameters and thermal noise of the four-terminal MOSFET in non-quasistatic operation," *Solid-State Electronics*, vol. 33, pp. 513–521, 1990.
154. L. Goldminz and Y. Nemirovsky, "Thermal noise in buried-channel MOSFET," *IEEE Transactions on Electron Devices*, vol. 39, pp. 2315–2332, October 1992.
155. R. Sarpeshkar, T. Delbrück, and C. A. Mead, "White noise in MOS transistors and resistors," *IEEE Circuits and Devices Magazine*, pp. 23–30, November 1993.
156. S. Tedja, J. Van der Spiegel, and H. H. Williams, "Analytical and experimental studies of thermal noise in MOSFET's," *IEEE Transactions on Electron Devices*, vol. 41, pp. 2069–2075, November 1994.
157. B. Wang, J. R. Hellums, and C. G. Sodini, "MOSFET thermal noise modeling for analog integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 833–835, July 1994.
158. F. Danneville, H. Happy, G. Dambrine, J.-M. Belquin, and A. Cappy, "Microscopic noise modeling and macroscopic noise models: how good a connection?," *IEEE Transactions on Electron Devices*, vol. 41, pp. 779–786, May 1994.
159. E. P. Vandamme, L. K. Vandamme, C. Claeys, E. Simoen, and R. J. Schreutelkamp, "Impact of silicidation on the excess noise behaviour of MOS transistors," *Solid-State Electronics*, vol. 38, pp. 1893–1897, November 1995.
160. G. Abowitz, E. Arnold, and E. Leventhal, "Surface states and $1/f$ noise in M.O.S. transistors," *IEEE Transactions on Electron Devices*, vol. ED-14, pp. 775–777, November 1967.
161. I. R. M. Mansour, R. J. Hawkins, and G. G. Bloodworth, "Measurement of current noise in MOS transistors from $5 \cdot 10^{-5}$ to 1 Hz," *The Radio and Electronic Engineer*, vol. 35, pp. 212–216, 1968.
162. E. A. Leventhal, "Derivation of $1/f$ noise in silicon inversion layers from carrier motion in a surface band," *Solid-State Electronics*, vol. 11, pp. 621–627, June 1968.
163. S. Christensson, I. Lundstrom, and C. Svensson, "Low frequency noise in MOS transistors—I. Theory," *Solid-State Electronics*, vol. 11, pp. 796–812, September 1968; "Low frequency noise in MOS transistors—II. Experiments," *ibid.*, pp. 813–820.
164. I. R. M. Mansour, R. J. Hawkins, and G. G. Bloodworth, "Physical model for the current noise spectrum of MOSTS," *British Journal of Applied Physics (Journal of Physics D: Applied Physics)* vol. 2, pp. 1063–1082, 1969.
165. H. E. Halladay and A. van der Ziel, "On the high frequency excess noise and equivalent circuit representation of the MOS-FET with n-type channel," *Solid-State Electronics*, vol. 12, pp 161–176, March 1969.
166. P. S. Rao, "The effect of the substrate upon the gate and drain noise parameters of MOSFETs," *Solid-State Electronics*, vol. 12, pp. 549–556, 1969.

167. L. D. Yau and A. van der Ziel, "Geometrical dependencies of the low-frequency generation-recombination noise in MOS transistors," *Solid-State Electronics*, vol. 12, pp. 903-905, 1969.
168. A. Takagi and A. van der Ziel, "Non-thermal noise in MOSFETs and MOS tetrodes," *Solid State Electronics*, vol. 12, p. 907, 1969.
169. N. R. Mantena and R. C. Lucas, "Experimental study of flicker noise in MIS field effect transistors," *Electronics Letters*, vol. 5, pp. 607-608, 1969.
170. F. Berz, "Theory of low frequency noise in Si MOSTs," *Solid-State Electronics*, vol. 13, pp. 631-647, 1970.
171. S. T. Hsu, "Surface state related $1/f$ noise in MOS transistors," *Solid-State Electronics*, vol. 13, pp. 1451-1459, 1970.
172. J. W. Hawkins and G. G. Bloodworth, "Two components of $1/f$ noise in MOS transistors," *Solid-State Electronics*, vol. 14, pp. 932-939, 1971.
173. F. M. Klaassen, "Characterization of low $1/f$ noise in MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-18, pp. 887-891, 1971.
174. J. W. Haslett and F. N. Trofimenkoff, "Effects of the substrate on surface state noise in silicon MOSFETs," *Solid-State Electronics*, vol. 15, pp. 117-131, 1972.
175. H. S. Fu and C. T. Sah, "Theory and experiments on surface $1/f$ noise," *IEEE Transactions on Electron Devices*, vol. ED-19, pp. 273-285, 1972.
176. R. S. Ronen, "Low-frequency $1/f$ noise in MOSFETs," *RCA Review*, vol. 34, pp. 280-307, 1973.
177. S. T. Hsu and A. van der Ziel, "A new type of flicker noise in microwave MOSFETs," *Solid-State Electronics*, vol. 18, pp. 885-886, 1975.
178. H. Katto, Y. Kamigaki, and Y. Itoh, "MOSFETs with reduced low frequency $1/f$ noise," *Japanese Journal of Applied Physics*, vol. 44, pp. 243-248, 1975.
179. S. T. Hsu, "Trapping noise in SOSMOST's," *Fourth International Conference on Physical Aspects of Noise in Solid State Devices*, Noordwijkerhout, The Netherlands, 1975.
180. A. van der Ziel, "Limiting flicker noise in MOSFETs," *Solid-State Electronics*, vol. 18, p. 1031, 1975.
181. F. N. Hooge, " $1/f$ noise," *Physica*, vol. 83B, pp. 14-23, 1976.
182. A. van der Ziel, "Dependence of flicker noise in MOSFET's on geometry," *Solid-State Electronics*, vol. 20, p. 267, 1977.
183. K. L. Wang, "Measurements of residual defects and $1/f$ noise in ion-implanted p-channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-25, pp. 478-484, 1978.
184. A. van der Ziel, "Some general relationships for flicker noise in MOSFETs," *Solid-State Electronics*, vol. 21, pp. 623-624, 1978.
185. R. P. Jindal and A. van der Ziel, "Carrier fluctuations noise in a MOSFET channel due to traps in the oxide," *Solid-State Electronics*, vol. 21, pp. 901-903, 1978.
186. P. Gentil, "Bruit bass fréquence du transistor MOS—1ere partie," *L'onde électrique*, vol. 58, pp. 565-575, August-September 1978; 2nd part, *ibid*, pp. 645-652, October 1978.
187. P. Victorovich and P. Gentil, "Influence of the depth of interface states in the insulator on the noise properties of MOS transistors," *Solid-State Electronics*, vol. 22, pp. 21-23, 1979.
188. W. V. Backensto and C. R. Viswanathan, "Bias-dependent $1/f$ noise model of an m.o.s. transistor," *IEE Proceedings*, vol. 127, part I, pp. 87-93, April 1980.
189. L. K. J. Vandamme and H. M. M. de Werd, " $1/f$ noise model for MOST's biased in nonohmic region," *Solid-State Electronics*, vol. 23, pp. 325-329, 1980.
190. L. K. J. Vandamme, "Model for $1/f$ noise in MOS transistors biased in the linear region," *Solid-State Electronics*, vol. 23, pp. 317-323, 1980.
191. P. Gentil and A. Mounib, "Equivalent input spectrum and drain current spectrum for $1/f$ noise in short channel MOS transistors," *Solid-State Electronics*, vol. 24, pp. 411-414, 1981.
192. H. S. Park, A. van der Ziel, and S. T. Liu, "Comparison of two $1/f$ noise models in MOSFET's," *Solid-State Electronics*, vol. 23, pp. 213-217, 1982.
193. H. Mikoshiba, " $1/f$ noise in n-channel silicon-gate MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 965-970, June 1982.
194. G. Reimbold, "Modified $1/f$ trapping noise theory and experiments in MOS transistors biased from weak to strong inversion—influence of interface states," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 1190-1198, September 1984.

195. J. M. Pimbley and G. Gildenblat, "Effect of hot-electron stress on low frequency MOSFET noise," *IEEE Electron Device Letters*, vol. EDL-5, pp. 345–347, September 1984.
196. Z. Celik and T. Y. Hsiang, "Study of $1/f$ noise in N-MOSFETs: linear region," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 2797–2802, December 1985.
197. A. van der Ziel, "Integral expression for $1/f$ noise in MOSFETs at arbitrary drain bias," *Solid-State Electronics*, vol. 29, pp. 29–30, 1986.
198. T. Watanabe, "Low-noise operation in buried-channel MOSFET's," *IEEE Electron Device Letters*, vol. EDL-6, pp. 317–319, July 1985.
199. A. van der Ziel, "Reconciliation of Klaassen's and Reimbold's theories of $1/f$ noise in MOSFETs," *Solid-State Electronics*, vol. 29, pp. 967–968, 1986.
200. B. Pellegrini, "On mobility-fluctuation origin on $1/f$ noise," *Solid-State Electronics*, vol. 29, pp. 1279–1287, 1986.
201. Z. C.-Butler and T. Y. Hsiang, "Spectral dependence of $1/f$ noise on gate bias in N-MOSFETs," *Solid-State Electronics*, vol. 30, pp. 419–423, 1987.
202. G. Ghibaudo, "A simple derivation of Reimbold's drain current spectrum formula for flicker noise in MOSFETs," *Solid-State Electronics*, vol. 30, pp. 1037–1038, 1987.
203. N. Mutoh and N. Teranishi, "New empirical relation form MOSFET $1/f$ noise unified over linear and saturation regions," *Solid-State Electronics*, vol. 31, pp. 1675–1680, 1988.
204. C. Surya and T. Y. Hsiang, "Surface mobility fluctuations in metal-oxide-semiconductor field-effect transistors," *Physical Review B*, vol. 35, p. 6342, 1987.
205. F. Grabowski, "Influence of dynamical interactions between density and mobility of carriers in the channel on $1/f$ noise of MOS transistors below saturation—I. Mechanisms," *Solid-State Electronics*, vol. 32, pp. 909–913, 1989.
206. F. Grabowski, "Influence of dynamical interactions between density and mobility of carriers in the channel of $1/f$ noise of MOS transistors below saturation—II. Implications," *Solid-State Electronics*, vol. 32, pp. 915–918, 1989.
207. G. Ghibaudo, "On the theory of carrier number fluctuations in MOS devices," *Solid-State Electronics*, vol. 32, pp. 563–565, 1989.
208. K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors," *IEEE Transactions on Electron Devices*, vol. 37, pp. 654–665, March 1990.
209. T. G. M. Kleinpenning, "On $1/f$ trapping noise in MOST's," *IEEE Transactions on Electron Devices*, vol. 37, pp. 2084–2089, September 1990.
210. Z.-H. Fang, A. Chovet, Q.-P. Zhu, and J.-N. Zhao, "Theory and applications of $1/f$ trapping noise in MOSFETs for the whole biasing ranges," *Solid-State Electronics*, vol. 34, pp. 327–333, 1991.
211. S.-L. Jang, "Analytical low-frequency $1/f$ noise model for lightly-doped-drain MOSFETs operating in the linear region," *Solid-State Electronics*, vol. 36, pp. 899–903, 1993.
212. S.-L. Jang and P.-C. Chang, "Low-frequency noise characteristics of lightly doped-drain MOSFETs," *Solid-State Electronics*, vol. 36, pp. 1007–1010, 1993.
213. L. K. J. Vandamme, X. Li, and D. Rigaud, " $1/f$ noise in MOS devices, mobility or number fluctuations?," *IEEE Transactions on Electron Devices*, vol. 41, November 1994.
214. J. H. Scofield, N. Borland, and D. M. Fleetwood, "Reconciliation of different gate-voltage dependencies of $1/f$ noise in n-MOS and p-MOS transistors," *IEEE Transactions on Electron Devices*, vol. 41, pp. 1946–1952, November 1994.
215. D. M. Fleetwood, T. L. Meisenheimer, and J. H. Scofield, " $1/f$ noise and radiation effects in MOS devices," *IEEE Transactions on Electron Devices*, vol. 41, p. 1953, November 1994.
216. J. Chang, A. A. Abidi, and C. R. Viswanathan, "Flicker noise in CMOS transistors from subthreshold to strong inversion at various temperatures," *IEEE Transactions on Electron Devices*, vol. 41, pp. 1965–1971, November 1994.
217. D. R. Wolters, A. T. A. Zegers-van Duijnhoven, and R. Augur, "Model for $1/f$ noise in MOSFETs and interconnects," *IEEE International Electron Devices Meeting*, pp. 177–180, December 1994.
218. S.-L. Jang, Comments on "Analytical low-frequency $1/f$ noise model for lightly-doped-drain MOSFETs operating in the linear region," *Solid-State Electronics*, vol. 37, pp. 1903–1904, November 1994.

219. X. Li, C. Barros, E. P. Vandamme, and K. L. Vandamme, "Parameter extraction and $1/f$ noise in a surface and a bulk-type, p -channel LDD MOSFET," *Solid-State Electronics*, vol. 37, pp. 1853–1862, November 1994.
220. C. Hu, J. Zhao, G. P. Li, P. Liu, E. Worley, J. White, and R. Kjar, "The effects of plasma etching induced gate oxide degradation on MOSFET's $1/f$ noise," *IEEE Electron Device Letters*, vol. 16, pp. 61–63, February 1995.
221. C. Hu, G. P. Li, E. Worley, and J. White, "Consideration of low-frequency noise in MOSFET's for analog performance," *IEEE Electron Device Letters*, vol. 17, pp. 552–554, December 1996.
222. P. Morfouli, G. Ghibaudo, T. Ouisse, E. Vogel, W. Hill, V. Misra, P. McLarty, and J. J. Wortman, "Low-frequency noise characterization of n - and p -MOSFET's with ultrathin oxynitride gate films," *IEEE Electron Device Letters*, vol. 17, pp. 395–397, August 1996.
223. M. Stegherr, "Flicker noise in hot electron degraded short channel MOSFETs," *Solid-State Electronics*, vol. 27, pp. 1055–1056, 1984.
224. R. P. Jindal, "Noise phenomena in submicron channel length nMOS transistors," *Noise in Physical Systems and $1/f$ Noise*, 1985 Elsevier Science B.V., The Netherlands, 1986.
225. Z. H. Fang, S. Cristoloveanu, and A. Chovet, "Analysis of hot-carrier-induced aging from $1/f$ noise in short-channel MOSFET's," *IEEE Electron Device Letters*, vol. EDL-7, pp. 371–373, June 1986.
226. B. Boukriss, H. Haddara, S. Cristoloveanu, and A. Chovet, "Modeling of the $1/f$ noise overshoot in short-channel MOSFET's locally degraded by hot-carrier injection," *IEEE Electron Device Letters*, vol. 10, pp. 433–436, October 1989.
227. A. Comeau, Comments on "Modeling of the $1/f$ noise overshoot in short-channel MOSFET's locally degraded by hot-carrier injection," *IEEE Electron Device Letters*, vol. 11, p. 129, March 1990.
228. M.-H. Tsai and T.-P. Ma, " $1/f$ noise in hot-carrier damaged MOSFET's: Effects of oxide charge and interface traps," *IEEE Electron Device Letters*, vol. 14, pp. 256–258, May 1993.
229. C.-H. Cheng and C. Surya, "The effect of hot-electron injection on the properties of flicker noise in N -channel MOSFETs," *Solid-State Electronics*, vol. 36, pp. 475–479, 1993.
230. D. Rigaud and M. Valenza, "Correlation measurement of carrier multiplication noise sources in MOS transistors at low frequencies," *IEEE Transactions on Electron Devices*, vol. 41, pp. 2076–2081, November 1994.
231. D. P. Triantis, A. N. Birbas, and D. Kondis, "Thermal noise modeling for short-channel MOSFET's," *IEEE Transactions on Electron Devices*, vol. 43, pp. 1950–1955, November 1996.
232. M. Aoki and M. Kato, "Hole-induced $1/f$ noise increase in MOS transistors," *IEEE Electron Device Letters*, vol. 17, pp. 118–120, March 1996.
233. D. P. Triantis, A. N. Birbas, and S. E. Plevridis, "Induced gate noise in MOSFET revisited: The submicron case," *Solid-State Electronics*, vol. 41, pp. 1937–1942, 1997.
234. K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant, "Discrete resistance switching in submicrometer silicon inversion layers: Individual interface traps and low-frequency ($1/f$?) noise," *Physical Review Letters*, vol. 52, p. 228, 1984.
235. M. J. Uren, D. J. Day, and M. J. Kirton, " $1/f$ and random telegraph noise in silicon metal-oxide-semiconductor field-effect transistors," *Applied Physics Letters*, vol. 47, p. 1195, 1985.
236. K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "Random telegraph noise of deep-submicrometer MOSFET's," *IEEE Electron Device Letters*, vol. 11, pp. 90–92, February 1990.
237. P. Fang, K. K. Hung, P. K. Ko, and C. Hu, "Hot-electron-induced traps studied through the random telegraph noise," *IEEE Electron Device Letters*, vol. 12, pp. 273–275, June 1991.
238. O. R. D. Buisson, G. Ghibaudo, and J. Brini, "Model for drain current RTS amplitude in small-area MOS transistors," *Solid-State Electronics*, vol. 35, pp. 1273–1276, 1992.
239. E. Simoen, B. Dierickx, C. L. Claeys, and G. J. Declerck, "Explaining the amplitude of RTS noise in submicrometer MOSFET's," *IEEE Transactions on Electron Devices*, vol. 39, pp. 422–429, February 1992.
240. M.-H. Tsai and T.-P. Ma, "The impact of device scaling on the current fluctuations in MOSFET's," *IEEE Transaction on Electron Devices*, vol. 41, pp. 2061–2068, November 1994.
241. Z. Shi, J.-P. Miéville, and M. Dutoit, "Random telegraph signals in deep submicron n -MOSFET's," *IEEE Transactions on Electron Devices*, vol. 41, pp. 1161–1168, July 1994.
242. Y. Tsiividis, K. Suyama, and K. Vavelidis, "A simple 'Reconciliation' MOSFET model valid in all regions," *Electronics Letters*, March 1995.

PROBLEMS

- 8.1. Restate the definitions of g_m , g_{mb} , and g_{sd} in terms of voltages V_G , V_B , V_D , and V_S , all taken with respect to an arbitrary reference. Put these in a form as simple as possible.
- 8.2. Discuss what types of I - V plots you would need in order to determine g_m , g_{mb} , and g_{sd} graphically, and describe the procedure for determining these parameters. Discuss qualitatively the accuracy you would expect from such a procedure.
- 8.3. Prove (8.2.11).
- 8.4. For an n -channel device in strong inversion with $V_{T0} = 0.5$ V, $V_{SB} = 0$ V, $(W/L)\mu C'_{ox} = 30 \mu\text{A/V}^2$, and $\alpha = 1.5$, plot: (a) g_m vs. V_{GS} with V_{DS} as a parameter, (b) g_m vs. V_{DS} with V_{GS} as a parameter.
- 8.5. Prove (8.2.16) and (8.2.17).
- 8.6. Assume that the effective mobility varies with V_{GS} according to (4.10.20), with $\theta_B = 0$. Show how (8.2.16) should be modified to take this effect into account.
- 8.7. Prove (8.2.19) and (8.2.20).
- 8.8. For a given I_{DS} and V_{SB} , state qualitatively what will happen to the ratios g_{mb}/g_m and g_{sd}/g_m if: (a) the substrate doping is increased; (b) the channel length is increased.
- 8.9. Prove (8.2.30) and (8.2.31).
- 8.10. Prove (8.2.35).
- 8.11. Show that an appropriate "charge sharing" model (Sec. 6.3) can predict an increase of g_{sd} in saturation with increasing drain junction depth, other things being equal.
- 8.12. Prove (8.2.40).
- 8.13. Prove (8.2.41) to (8.2.44).
- 8.14. Prove (8.2.49) and (8.2.50).
- 8.15. Using the equivalent small-signal circuit of Fig. 8.17, perform all "measurements" illustrated in Figs. 8.2 and Fig. 8.15b, c, and d. Show that in each case the elements do not interfere with each other, i.e., the results given next to each figure are still obtained.
- 8.16. Prove (8.3.8) to (8.3.12) using the simplifying assumptions stated above them.
- 8.17. Show that, for the device of Fig. 8.19, assuming V_D and V_S are almost equal, a very small change ΔV of V_S only will decrease the gate charge by $\frac{1}{2} C_{ox} \Delta V$.
- 8.18. Derive the results in (8.3.24) to (8.3.28) by differentiating the saturation charge expressions (7.4.26) and (7.4.30), using the simplifying assumptions made in Sec. 8.3 (i.e., that $\alpha = \alpha_1$ and that the variation of α_1 with V_S and V_B is negligible).
- 8.19. Show that (7.4.21) and (7.4.25) *cannot* be used in the capacitance definitions to derive (8.3.16) and (8.3.17). Explain the reasons for this.
- 8.20. For the device of Prob. 8.4 operating in strong inversion, plot all capacitances (a) versus V_{GS} , with V_{DS} as a parameter, and (b) versus V_{DS} , with V_{GS} as a parameter.
- 8.21. Prove (8.3.32).
- 8.22. Prove (8.3.35). [Hint: Use (4.6.2), (4.6.6), (4.6.7), and the fact that, at the upper limit of weak inversion, $\psi_s \approx 2\phi_F + V_{SB}$ (Sec. 3.4.1).]
- 8.23. Show that the models of Fig. 8.32b and c are equivalent.
- 8.24. Derive (8.5.14), assuming Δv is very small.
- 8.25. (a) For a transistor in saturation, with $\mu = 600 \text{ cm}^2/(\text{V}\cdot\text{s})$, $t_{ox} = 80 \text{ \AA}$, $\alpha = 1.2$, $W = L = 2 \mu\text{m}$, $V_T = 0.5$ V, and $V_{GS} = 2$ V, evaluate S_{iw} , S_{vw} , and R_n at $T = 200, 300$, and 400 K.
(b) Find the mean square value of the thermal noise current for the above device at 300 K, in the band from 100 to 200 kHz.

- 8.26. The MOS transistor is sometimes modeled by a noiseless transistor with a noise voltage source in series with its gate, representing the equivalent input noise voltage. If capacitances are to be added to this model, on which side of the above source should C_{gs} , C_{gd} , and C_{gb} be connected? Show that, for one of these choices, a problem will result at high frequencies if the device is driven by a signal source (between gate and source) with a significant internal resistance.
- 8.27. Prove (8.5.27) and (8.5.28) as well as the claims in the footnote following (8.5.28).
- 8.28. A transistor is operating in weak inversion at $T = 300$ K, with $I'_{DS} = 10$ nA, and $n = 1.5$. Plot S_{iw} versus V_{DS} .
- 8.29. The device of Prob. 8.25, operating at $T = 300$ K, exhibits a "corner frequency" f_c (see Fig. 8.33) of 4 kHz. Assuming $c = 1$ in (8.5.29) and $k_1 = 7 \times 10^{-31}$ C² · cm⁻², calculate the mean square value of the total current noise (flicker plus thermal) in the band of 20 Hz to 200 kHz.
- 8.30. Develop an expression for an equivalent input noise resistance representing both thermal and flicker noise in strong inversion.

CHAPTER 9

HIGH- FREQUENCY SMALL-SIGNAL MODELS

9.1 INTRODUCTION

In this chapter we study models that are valid in a wider frequency range than the five-capacitance quasi-static model of Chap. 8.¹⁻⁵⁵ The first model considered is what will be called the *complete quasi-static* model. This model represents an attempt to take every possible advantage of the quasi-static assumption, and gives an improved upper frequency limit of validity. Beyond that limit, though, the model becomes very inaccurate. It is then necessary to consider *non-quasi-static* models. Such models will be considered after a general discussion of y-parameter models. A comparison of all models presented will be performed. Noise at high frequencies will then be considered. The chapter will conclude with a discussion of modeling for radio-frequency (RF) applications, including the effects of extrinsic elements.

All effects considered in this chapter will be understood to be for the *intrinsic* part of the transistor (see Fig. 7.1) except in Secs. 9.3 and 9.6. No impact ionization is assumed to be present.

9.2 A COMPLETE QUASI-STATIC MODEL

9.2.1 Complete Description of Capacitance Effects

In Sec. 8.3, we assumed quasi-static operation (defined in Sec. 7.2), and we modeled the capacitance effect of the drain, source, and substrate on the gate, and the effects

of the drain and source on the substrate. Clearly, we did not consider all possible combinations in choosing the above five effects. Nevertheless, the model was claimed to be satisfactory for many applications up to a certain frequency. We will now undertake the rigorous development of a *complete* quasi-static model. The model will be complete in the sense that the capacitance effect of *every* terminal on *every* other will be modeled.^{18,28-37,39,47,53}

Consider an intrinsic transistor with time-varying voltages, as shown in Fig. 9.1a. Lowercase letters with capital subscripts denote *total* quantities (as opposed to bias or small-signal quantities). In Sec. 7.3 we showed how the transport and charging components of the drain and source currents in quasi-static operation can be evaluated. The meaning of the drain and source charging components is rather subtle, and was discussed at length. The gate and substrate currents consist only of charging components. The four charging currents satisfy (7.3.15) and are given by (7.3.16a) to (7.3.16d). We now assume that the total voltages consist of a dc bias part and a small-signal part, as shown in Fig. 9.1b. Bias quantities will be represented by capital symbols with capital subscripts; small-signal quantities will be represented by lowercase symbols with lowercase subscripts. If the small signals $v_d(t) = v_g(t) = v_b(t) = v_s(t) = 0$, for all t , then $dv_D/dt = dv_G/dt = dv_B/dt = dv_S/dt = 0$, and thus all currents in (7.3.16) are zero. If the small-signal voltages are nonzero but vary sufficiently slowly,

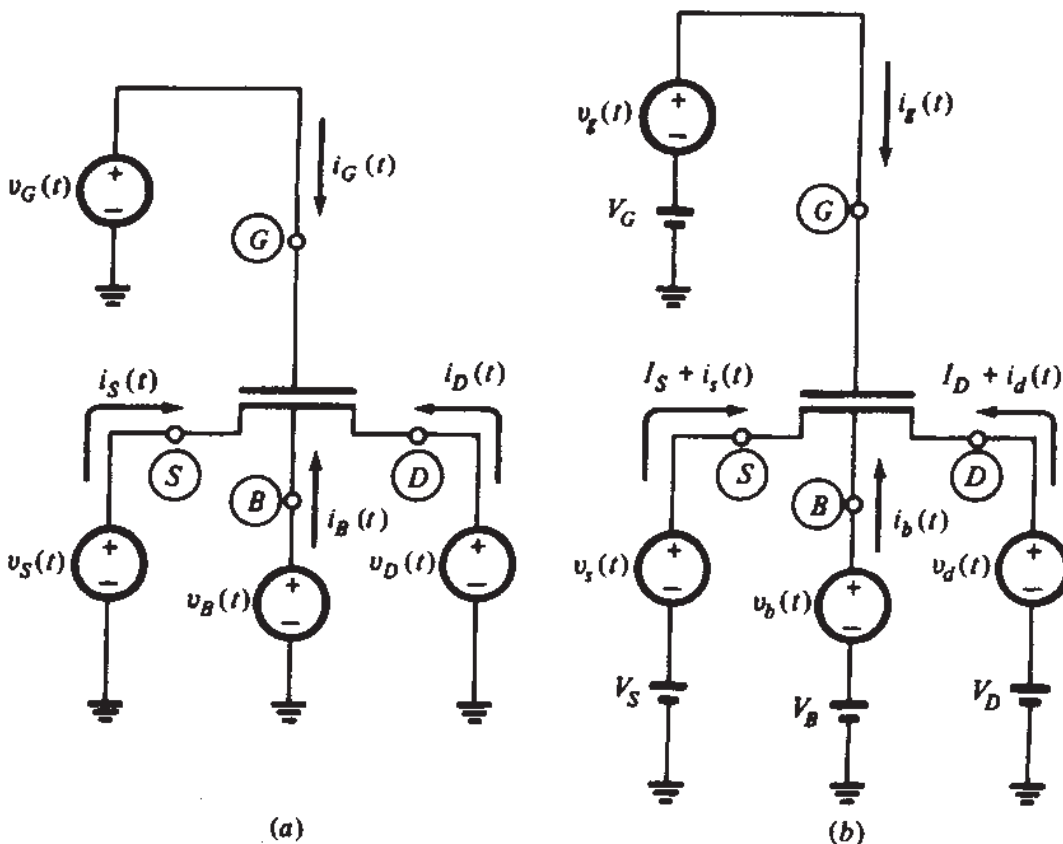


FIGURE 9.1

(a) A transistor with four time-varying terminal voltages; (b) a transistor with terminal voltages consisting of a dc bias plus a time-varying small signal.

then all charging currents in (7.3.16) will be very small. They will be represented by $i_{da}(t)$, $i_g(t)$, $i_b(t)$, and $i_{sa}(t)$. Over the small range of variation of the small-signal voltages, the slopes of the form $\partial q_k / \partial v_L$, where q_k is any of the four charges and v_L any of the voltages in (7.3.16), will be assumed constant and equal to their values at the bias (or "operating") point ($v_D = V_D$, $v_G = V_G$, $v_B = V_B$, $v_S = V_S$). Denoting evaluation at this point by o , we define

$$C_{kk} = + \left. \frac{\partial q_K}{\partial v_K} \right|_o \quad (9.2.1a)$$

$$C_{kl} = - \left. \frac{\partial q_K}{\partial v_L} \right|_o, \quad l \neq k \quad (9.2.1b)$$

The choice of the algebraic signs in the above definitions is common. It will prove convenient in developing small-signal equivalent circuits later in this section, and in relating these circuits to the ones in Sec. 8.3. The sign choice in (9.2.1b) is consistent with (8.3.1) to (8.3.5), the negative sign of which was discussed above (8.3.1) (see also Prob. 9.1).

Of the capacitance parameters defined above, five (C_{gs} , C_{bs} , C_{gd} , C_{bd} , and C_{gb}) have precisely the meaning discussed in Sec. 8.3. Ways to measure C_{kk} and C_{kl} will be considered in the next section. Using the above definitions, we have, from (7.3.16) and the above discussion, the following expressions for the small-signal charging currents:

$$i_{da}(t) = +C_{dd} \frac{dv_d}{dt} - C_{dg} \frac{dv_g}{dt} - C_{db} \frac{dv_b}{dt} - C_{ds} \frac{dv_s}{dt} \quad (9.2.2a)$$

$$i_g(t) = -C_{gd} \frac{dv_d}{dt} + C_{gg} \frac{dv_g}{dt} - C_{gb} \frac{dv_b}{dt} - C_{gs} \frac{dv_s}{dt} \quad (9.2.2b)$$

$$i_b(t) = -C_{bd} \frac{dv_d}{dt} - C_{bg} \frac{dv_g}{dt} + C_{bb} \frac{dv_b}{dt} - C_{bs} \frac{dv_s}{dt} \quad (9.2.2c)$$

$$i_{sa}(t) = -C_{sd} \frac{dv_d}{dt} - C_{sg} \frac{dv_g}{dt} - C_{sb} \frac{dv_b}{dt} + C_{ss} \frac{dv_s}{dt} \quad (9.2.2d)$$

It will be helpful in this discussion not to associate the various capacitance parameters above with any physical capacitor-like structures in the MOS transistor. It is better to consider them for the present as simply quantities defined in the precise manner of (9.2.1). We note here that in general $C_{kl} \neq C_{lk}$. For example, consider a long-channel device in saturation. Varying the voltage at the drain will not affect the rest of the device because of pinchoff (assuming no channel length modulation). Hence, the gate charge will not change [see (7.4.30)], and, from (9.2.1b), C_{gd} will be zero. However, varying the gate voltage will change the inversion layer charge. As explained in Sec. 7.3, this change will be accomplished in part by the drain current temporarily becoming different from the transport value. The difference under small-

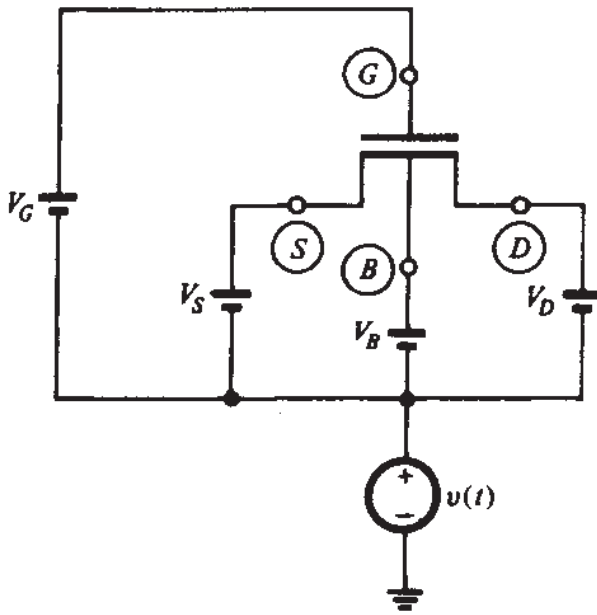


FIGURE 9.2

A transistor with all four terminal-to-ground small-signal voltages equal.

signal conditions is $i_{da}(t)$ and (assuming all other voltages are kept constant) is equal to $-C_{dg}(dv_g/dt)$ from (9.2.2a), which can only be nonzero if $C_{dg} \neq 0$. Another way to see this is to observe that the charge “associated with the drain”[†] *does* depend on the gate voltage even in saturation [see (7.4.28)]. Thus, from (9.2.1b), $C_{dg} \neq 0$. Therefore, it is seen that $C_{gd} \neq C_{dg}$, and *this can be verified by measurements*, as will be seen. This fact may seem strange at first, because we may have a tendency to think of C_{gd} and C_{dg} as the capacitances of two-terminal capacitors between gate and drain. However, such an interpretation is *not* correct. C_{gd} represents the effect of the drain on the gate, and C_{dg} represents the effect of the gate on the drain, in terms of charging currents. There is no reason to expect that the two effects are the same in general, just as there is no reason to expect that, at dc, the effect of the drain on the gate current (which is zero assuming no leakage) is the same as the effect of the gate on the drain current (which can be large). We will expand on this point later on.

We now make some important observations about the capacitance parameters and (9.2.2). First, assume that $v_d(t) = v_g(t) = v_b(t) = v_s(t) = v(t)$ in Fig. 9.1b. This is equivalent to the situation shown in Fig. 9.2. From (9.2.2a) we will have

$$i_{da}(t) = (C_{dd} - C_{dg} - C_{db} - C_{ds}) \frac{dv}{dt} \quad (9.2.3)$$

However, since there is no small-signal voltage across any two of the terminals in Fig. 9.2, all terminal small-signal currents must be zero. Since this must be true even for nonzero dv/dt , (9.2.3) implies that

$$C_{dd} - C_{dg} - C_{db} - C_{ds} = 0 \quad (9.2.4)$$

[†]To appreciate the points made here requires a careful interpretation of the term “charge associated with the drain.” This term can easily be misinterpreted. A related extensive discussion, which is very relevant to our present topic, has been given in Sec. 7.3.

Let us now make another observation. As follows from (7.3.15), the small-signal charging currents must add up to zero:

$$i_{da}(t) + i_g(t) + i_b(t) + i_{sa}(t) = 0 \quad (9.2.5)$$

Consider a case in which $dv_g/dt = dv_b/dt = dv_s/dt = 0$. Using this in (9.2.2a) to (9.2.2d) and the results in (9.2.5), we have

$$(C_{dd} - C_{gd} - C_{bd} - C_{sd}) \frac{dv_d}{dt} = 0 \quad (9.2.6)$$

which, since it must be valid even for nonzero dv_d/dt , implies that

$$C_{dd} - C_{gd} - C_{bd} - C_{sd} = 0 \quad (9.2.7)$$

Equations (9.2.4) and (9.2.7) provide two expressions for C_{dd} in terms of other capacitance parameters. Similar expressions can be derived in the same manner for C_{gg} , C_{bb} , and C_{ss} . Thus, we have

$$C_{dd} = C_{dg} + C_{db} + C_{ds} = C_{gd} + C_{bd} + C_{sd} \quad (9.2.8a)$$

$$C_{gg} = C_{gd} + C_{gb} + C_{gs} = C_{dg} + C_{bg} + C_{sg} \quad (9.2.8b)$$

$$C_{bb} = C_{bd} + C_{bg} + C_{bs} = C_{db} + C_{gb} + C_{sb} \quad (9.2.8c)$$

$$C_{ss} = C_{sd} + C_{sg} + C_{sb} = C_{ds} + C_{gs} + C_{bs} \quad (9.2.8d)$$

An interesting result that can be derived from these equations is considered in Prob. 9.2.

Our next observation is simply that, if three of the small-signal charging currents are known, the fourth can be determined from (9.2.5). Thus, any one among the four equations (9.2.2) can be omitted without losing any information. We will omit the last equation from now on.

Without loss of generality, we can write (see Fig. 9.3)

$$v_D = v_{DS} + v_S \quad (9.2.9a)$$

$$v_G = v_{GS} + v_S \quad (9.2.9b)$$

$$v_B = v_{BS} + v_S \quad (9.2.9c)$$

Using analogous expressions for the small-signal voltages in (9.2.2a) gives:

$$i_{da}(t) = C_{dd} \frac{dv_{ds}}{dt} - C_{dg} \frac{dv_{gs}}{dt} - C_{db} \frac{dv_{bs}}{dt} + (C_{dd} - C_{dg} - C_{db} - C_{ds}) \frac{dv_s}{dt} \quad (9.2.10)$$

The quantity in parentheses is equal to zero, as can be seen from (9.2.8a). Thus,

$$i_{da}(t) = C_{dd} \frac{dv_{ds}}{dt} - C_{dg} \frac{dv_{gs}}{dt} - C_{db} \frac{dv_{bs}}{dt} \quad (9.2.11)$$

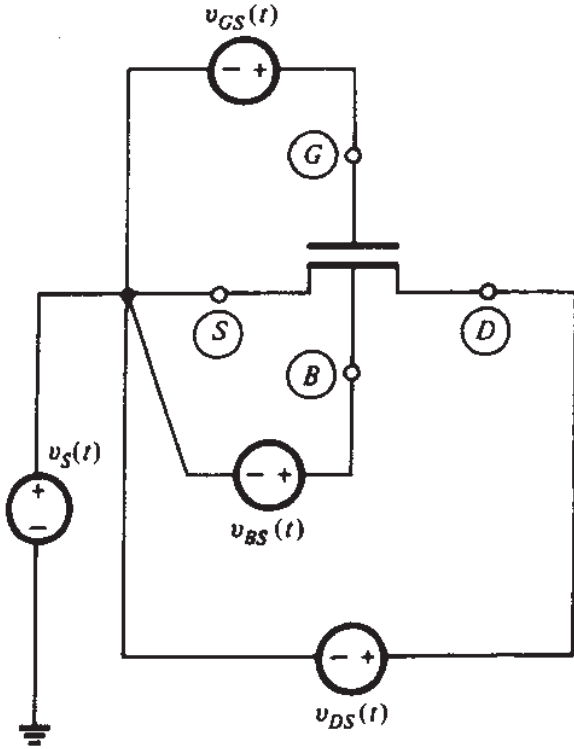


FIGURE 9.3
A transistor with terminal voltages referenced to the source.

Similar relations can be obtained from (9.2.2*b*) and (9.2.2*c*). Equation (9.2.2*d*) will not be considered anymore, as explained above. Thus, we finally obtain

$$i_{da}(t) = +C_{dd} \frac{dv_{ds}}{dt} - C_{dg} \frac{dv_{gs}}{dt} - C_{db} \frac{dv_{bs}}{dt} \quad (9.2.12a)$$

$$i_g(t) = -C_{gd} \frac{dv_{ds}}{dt} + C_{gg} \frac{dv_{gs}}{dt} - C_{gb} \frac{dv_{bs}}{dt} \quad (9.2.12b)$$

$$i_b(t) = -C_{bd} \frac{dv_{ds}}{dt} - C_{bg} \frac{dv_{gs}}{dt} + C_{bb} \frac{dv_{bs}}{dt} \quad (9.2.12c)$$

It is thus clear from the above discussion that a complete small-signal description of the charging mechanisms requires no less (and no more) than *nine* independent capacitance parameters.

9.2.2 Small-Signal Equivalent Circuit Topologies

Small-signal equivalent circuits can be derived to represent (9.2.12). Many such circuits can be constructed; the most straightforward one is shown in Fig. 9.4*a*. This circuit can be verified by writing Kirchhoff's current law for terminals *g*, *d*, and *b*, which results in (9.2.12). We do not yet have a complete model for the transistor though, because the transport component of the drain and source currents, originally in (7.3.4),

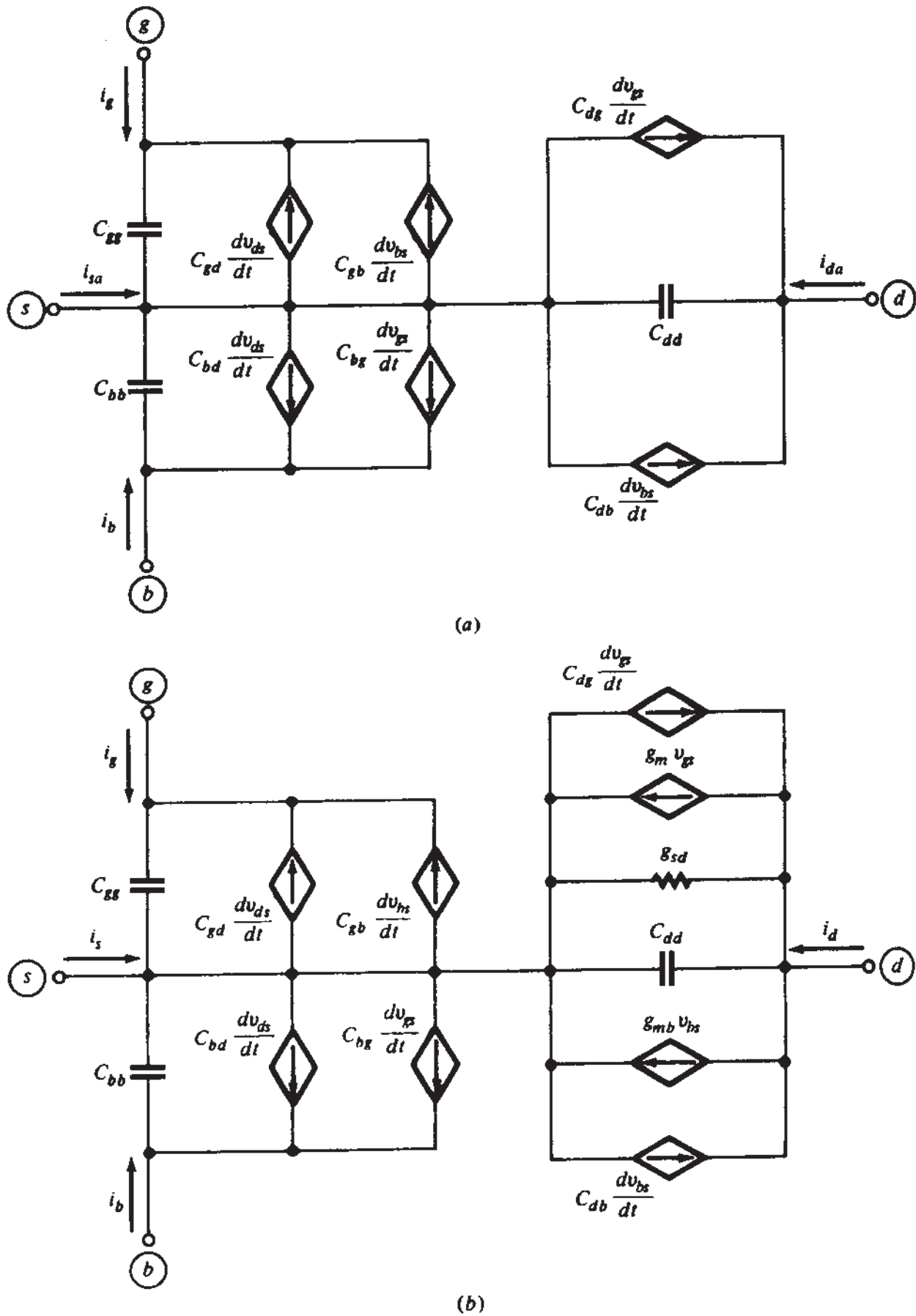


FIGURE 9.4

(a) A small-signal equivalent circuit for the charging current action; (b) a complete quasi-static model resulting from (a) by adding the transport current modeling elements of Fig. 8.3.

was omitted from the subsequent development. Now, if the small-signal voltages in Fig. 9.1b are zero, all terminal voltages and currents are fixed. The charging components are thus zero, and (7.3.4) gives (using capital I to represent dc current)

$$I_D = I_T \quad (9.2.13a)$$

$$I_S = -I_T \quad (9.2.13b)$$

In the general case, where the small-signal voltages are not zero, (7.3.4) leads to

$$I_D + i_d(t) = I_T + i_t(t) + i_{da}(t) \quad (9.2.14a)$$

$$I_S + i_s(t) = -I_T - i_t(t) + i_{sa}(t) \quad (9.2.14b)$$

where now both small-signal transport components $i_t(t)$ and small-signal charging components appear in the drain and source currents. Subtracting (9.2.13) from (9.2.14), we obtain equations relating only the small-signal currents:

$$i_d(t) = i_t(t) + i_{da}(t) \quad (9.2.15a)$$

$$i_s(t) = -i_t(t) + i_{sa}(t) \quad (9.2.15b)$$

The small-signal transport current has been modeled by the three-element combination of Fig. 8.3. Since this current *adds* to the charging current in (9.2.15a), the corresponding part of the model of Fig. 8.3 should be added in *parallel* with the rest of the elements between drain and source in Fig. 9.4a. This yields the complete model of Fig. 9.4b.

A different small-signal equivalent circuit will now be derived. Let us use

$$v_{ds} = v_{dg} + v_{gs} = -v_{gd} + v_{gs} \quad (9.2.16a)$$

$$v_{bs} = v_{bg} + v_{gs} = -v_{gb} + v_{gs} \quad (9.2.16b)$$

Substituting these equations in (9.2.12b), we obtain

$$\begin{aligned} i_g(t) &= -C_{gd} \left(-\frac{dv_{gd}}{dt} + \frac{dv_{gs}}{dt} \right) + C_{gg} \frac{dv_{gs}}{dt} - C_{gb} \left(-\frac{dv_{gb}}{dt} + \frac{dv_{gs}}{dt} \right) \\ &= C_{gd} \frac{dv_{gd}}{dt} + C_{gb} \frac{dv_{gb}}{dt} + (C_{gg} - C_{gd} - C_{gb}) \frac{dv_{gs}}{dt} \end{aligned} \quad (9.2.17)$$

Using (9.2.8b), we find that this becomes

$$i_g(t) = C_{gd} \frac{dv_{gd}}{dt} + C_{gb} \frac{dv_{gb}}{dt} + C_{gs} \frac{dv_{gs}}{dt} \quad (9.2.18)$$

Using similar manipulations in (9.2.12a) and (9.2.12c), we can write (9.2.12) in the form

$$i_{da}(t) = C_{gd} \frac{dv_{dg}}{dt} + C_{sd} \frac{dv_{ds}}{dt} + C_{bd} \frac{dv_{db}}{dt} - C_m \frac{dv_{gs}}{dt} - C_{mb} \frac{dv_{bs}}{dt} \quad (9.2.19a)$$

$$i_g(t) = C_{gd} \frac{dv_{gd}}{dt} + C_{gb} \frac{dv_{gb}}{dt} + C_{gs} \frac{dv_{gs}}{dt} \quad (9.2.19b)$$

$$i_b(t) = C_{bd} \frac{dv_{bd}}{dt} + C_{gb} \frac{dv_{bg}}{dt} - C_{mx} \frac{dv_{gb}}{dt} + C_{bs} \frac{dv_{bs}}{dt} \quad (9.2.19c)$$

where

$$C_m = C_{dg} - C_{gd} \quad (9.2.20a)$$

$$C_{mb} = C_{db} - C_{bd} \quad (9.2.20b)$$

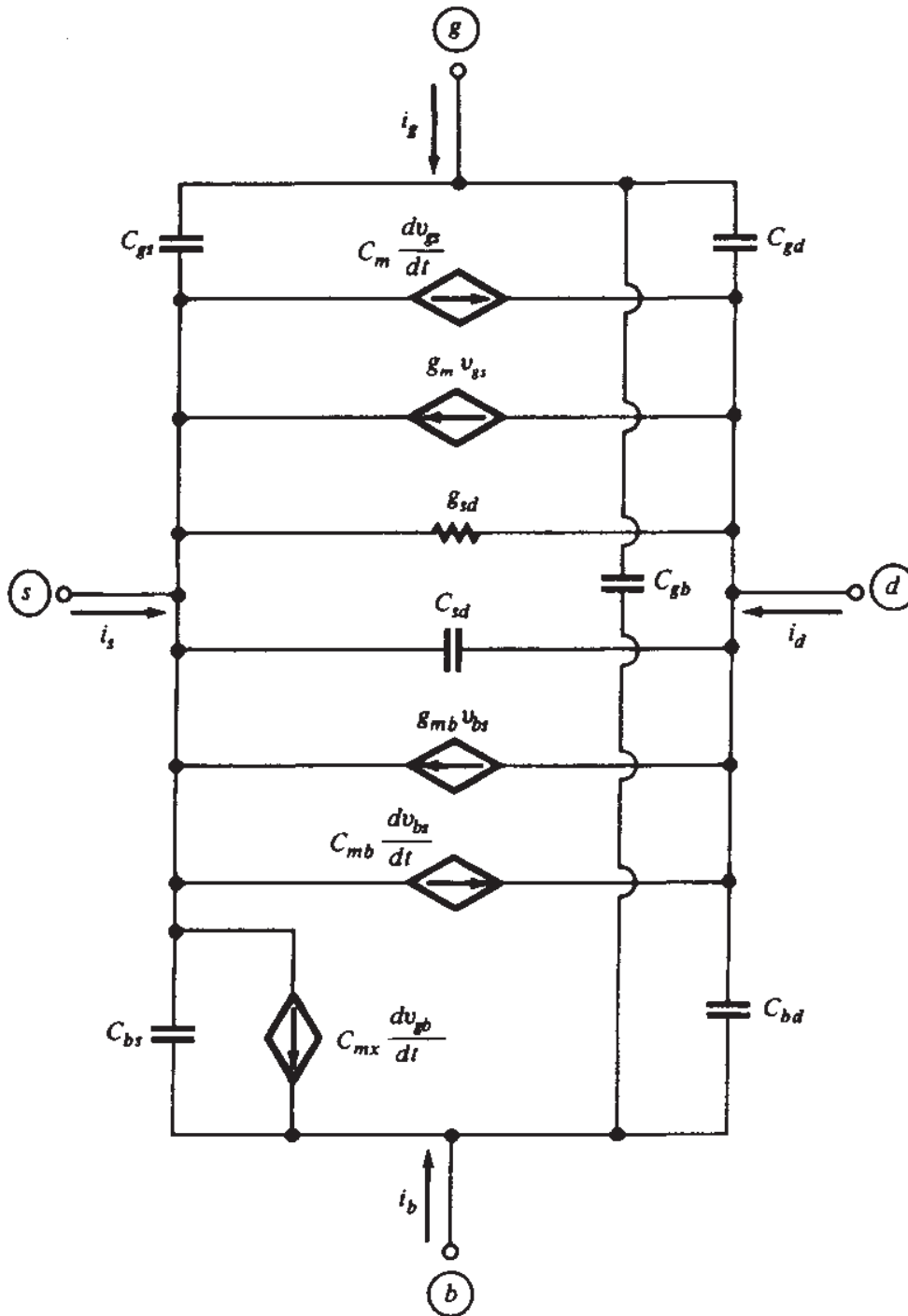
$$C_{mx} = C_{bg} - C_{gb} \quad (9.2.20c)$$

Equation (9.2.19b) has already been proved. To prove (9.2.19a) and (9.2.19c), simply express all voltages in these equations in terms of v_{ds} , v_{gs} , and v_{bs} and use (9.2.20) and (9.2.8); this will give (9.2.12a) and (9.2.12c).

Equations (9.2.19) are easily represented by an equivalent circuit. When the elements of Fig. 8.3 are added to it, we obtain the result shown in Fig. 9.5.³⁶ Five capacitors shown in this figure (C_{gs} , C_{gd} , C_{bs} , C_{bd} , and C_{gb}) are exactly the same as those in the model of Fig. 8.17. In other words, the model of Fig. 9.5 can be viewed as resulting from simply *augmenting* the model of Fig. 8.17 by four more elements, but *without* having to modify the elements already in that popular model, neither in meaning nor in value. This is an important property of the model in Fig. 9.5, in contrast to other models proposed in the literature. This point is considered further in Probs. 9.5 and 9.6.

Note that, in Fig. 9.5, at sufficiently low frequencies dv_{gs}/dt and dv_{bs}/dt will be small, and the currents proportional to these quantities can be neglected in comparison to the currents $g_m v_{gs}$ and $g_{mb} v_{bs}$, respectively. Similarly, the current through C_{sd} can then be neglected in comparison to the current through g_{sd} . Finally, we will find below that for the approximate strong-inversion model $C_{mx} = 0$. Thus, the model of Fig. 9.5 reduces to that of Fig. 8.17 at sufficiently low frequencies. More on model comparison will be found later in this chapter.

To provide more feeling for the model in Fig. 9.5, we will consider two experiments, as illustrated in Fig. 9.6. In Fig. 9.6a, a small-signal voltage is applied only at the drain, and the resulting small-signal current entering the gate is observed; in Fig. 9.6b, the opposite is done. The small-signal equivalent circuit used in each case on the right is that of Fig. 9.5, but only elements with nonzero current through them are shown for simplicity (short-circuited capacitances or resistances, and current sources proportional to zero voltages are omitted). As seen in *a*, the small-signal current en-

**FIGURE 9.5**

A complete quasi-static small-signal model. Although independently derived, the model can be viewed as resulting from the simpler model of Fig. 8.17 by adding four elements to it.

tering the gate is $-C_{gd}(dv_d/dt)$. Thus, C_{gd} actually represents the effect of terminal d on terminal g . However, although the capacitance C_{gd} is connected between g and d in the general case, it *does not* represent the total effect of g on d . This can be seen in Fig. 9.6b. Indeed, the current entering the drain is $g_m v_g - (C_{gd} + C_m)(dv_g/dt)$. Note that not only is there a conductive current here but also the capacitive current is *different* from that in *a*. Thus, $C_{gd} + C_m$ represents the capacitive effect of g on d . From

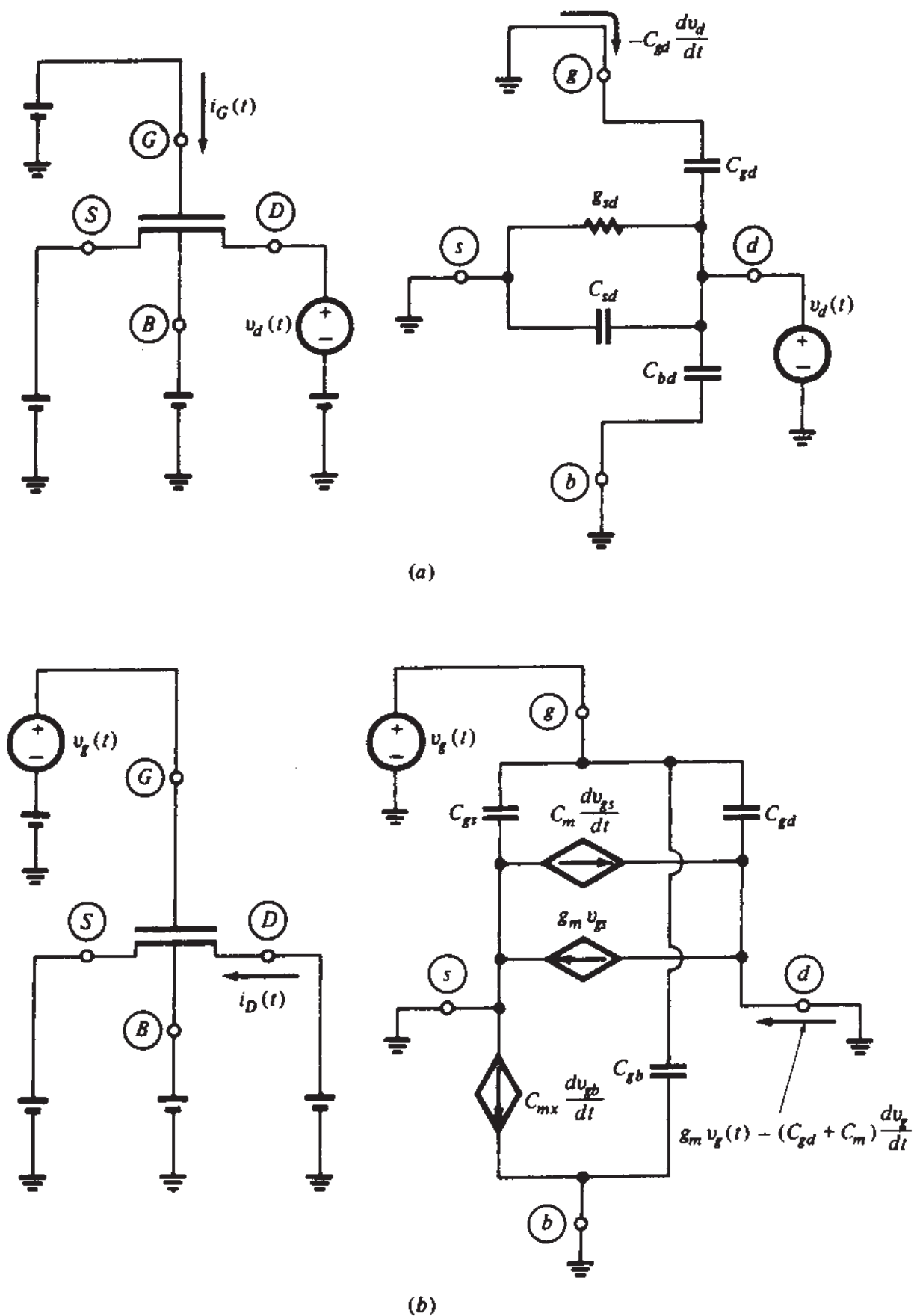


FIGURE 9.6

Illustration of the way in which the different effect of the gate and drain on each other is taken into account in the model of Fig. 9.5: (a) effect of drain voltage on gate circuit; (b) effect of gate voltage on drain current. In each part, the left half shows the complete circuit; the right half shows the small-signal equivalent circuit, in which elements with zero current through them are omitted.

(9.2.20a), $C_{gd} + C_m = C_{dg}$, which, in general, is *different* from C_{gd} , as we have already argued, and as we will see quantitatively in the following subsection. C_m is then a *transcapacitance*, taking care of the different effect of the gate and drain on each other in terms of charging currents, just as g_m is a transconductance taking care of the different effect of these two terminals on each other in terms of transport currents. Similar comments hold for C_{mb} and C_{mx} .

9.2.3 Evaluation of Capacitances

STRONG INVERSION. In Sec. 8.3, we used the charges corresponding to the approximate model to evaluate five capacitances: C_{gs} , C_{bs} , C_{gd} , C_{bd} , C_{gb} , and we gave several plots of these capacitances vs. V_{DS} and V_{GS} . The evaluation of the new capacitances that have been defined in this section is done in the same manner. Again, in deriving expressions for these capacitances we will use the simplifying assumptions of Sec. 8.3, that is, $\alpha = \alpha_1$ as given by (8.3.7), and $d\alpha_1/dV_s$ and $d\alpha_1/dV_B$ are negligible. C_{dg} is found by using $q_D = Q_D$ from (7.4.19) in the definition (9.2.1b).† This gives, after much algebra,

$$C_{dg} = C_{ox} \frac{4 + 28\eta + 22\eta^2 + 6\eta^3}{15(1 + \eta)^3} \quad (9.2.21)$$

with η as defined in (4.5.38) and plotted in Fig. 4.20. Similarly, we find:

$$C_{db} = (\alpha_1 - 1) C_{dg} \quad (9.2.22)$$

and, using (7.4.15),

$$C_{bg} = \frac{\alpha_1 - 1}{3\alpha_1} C_{ox} \left(\frac{1 - \eta}{1 + \eta} \right)^2 \quad (9.2.23)$$

Comparing the last equation to (8.3.12), we see that, for the approximate model with the simplifying assumptions we have made, $C_{bg} = C_{gb}$.

To evaluate C_{sd} , we use Q_s from (7.4.20) in the definition (9.2.1b) and obtain

$$C_{sd} = - \left[\frac{4}{15} C_{ox} \alpha_1 \frac{\eta + 3\eta^2 + \eta^3}{(1 + \eta)^3} \right] \quad (9.2.24)$$

†The charges and voltages in Sec. 7.4 were assumed to be dc quantities. As before, we will use the expressions developed in that section, with Q replaced by q and V replaced by v , under the assumption of quasi-static operation (Sec. 7.2).

Notice that this quantity is *negative* in nonsaturation (in saturation, C_{sd} becomes zero). This is in agreement with measurements.³² The negative value can be viewed intuitively as follows. Raising the drain voltage by an amount ΔV_D will increase the effective reverse bias at the drain end and will cause the magnitude of the inversion layer charge to decrease. Since Q_I is negative, this means a change $\Delta Q_I > 0$. This positive change in Q_I is shared by positive ΔQ_S and positive ΔQ_D . Hence, $C_{sd} = -\partial Q_S / \partial V_D$ will be negative.

Plots of C_{sd} , C_{dg} , C_{db} , and C_{bg} are given in Fig. 9.7. Comparisons to accurate results using the charge sheet model show that good accuracy is obtained except for C_{db} and C_{bg} when V_{SB} is small and V_{GS} and V_{DS} are large. The accuracy of C_{db} and C_{bg} can be improved as was done for C_{bs} and C_{bd} in Sec. 8.3, i.e., by replacing α_1 by a function of the form of (8.3.14).

The above relations, along with (8.3.8) to (8.3.12), give values for nine capacitances: C_{gs} , C_{bs} , C_{gd} , C_{dg} , C_{bd} , C_{db} , C_{gb} , C_{bg} , and C_{sd} . Any other capacitance parameter defined in this section can be found from these and (9.2.8) or (9.2.20). For example, using the latter we find

$$C_m = \frac{4}{15} C_{ox} \frac{1 + 2\eta - 2\eta^2 - \eta^3}{(1 + \eta)^3} \quad (9.2.25)$$

$$C_{mb} = (\alpha_1 - 1) C_m \quad (9.2.26)$$

$$C_{mx} = 0 \quad (9.2.27)$$

Thus, we have expressions for all nine capacitance parameters in the model of Fig. 9.5. Plots of C_m , C_{mb} , and C_{mx} vs. V_{DS} are given in Fig. 9.8. The accuracy for C_{mb} for small V_{SB} and large V_{GS} and V_{DS} is not good. It can be improved by replacing α_1 by a function of the form of (8.3.14). However, both this parameter and C_{db} are not of prime importance for most applications. Also, note that our simple model predicts $C_{bg} = C_{gb}$ or $C_{mx} = 0$. Accurate calculations³⁶ using the charge sheet model give, however,

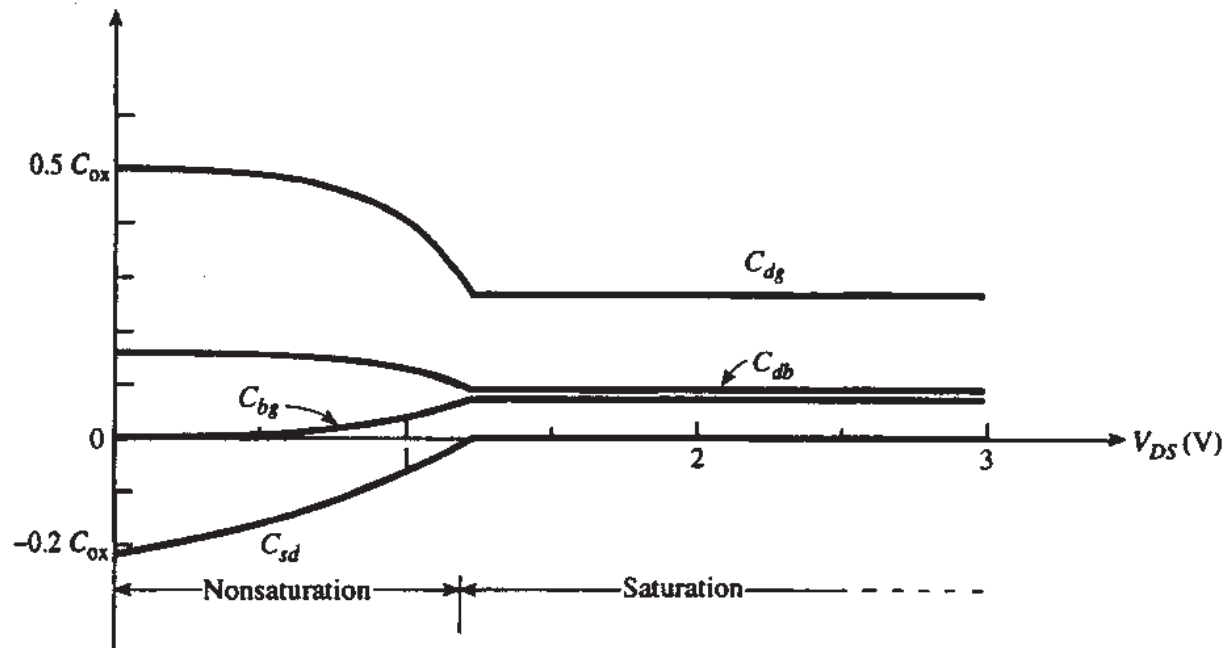
$$C_{bg} > C_{gb} \quad (9.2.28)$$

and, consequently,[†]

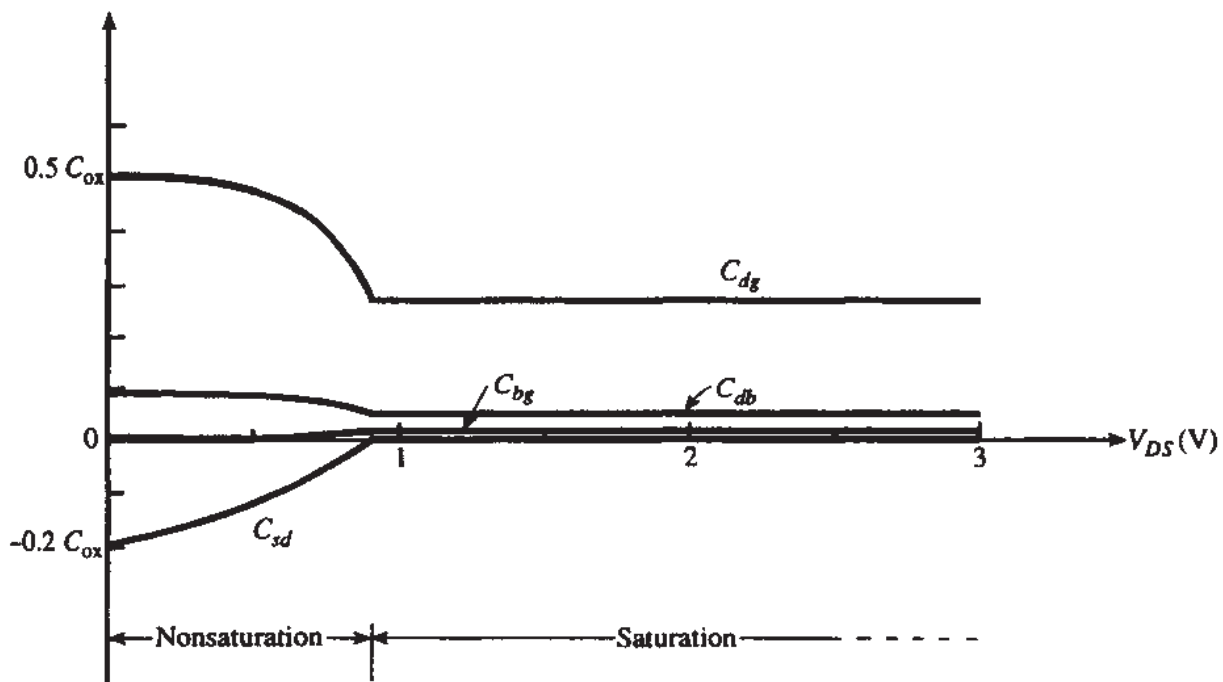
$$C_{mx} > 0 \quad (9.2.29)$$

Nevertheless, C_{mx} is very small and will be unimportant in most practical cases.

[†]The prediction that $C_{gb} = C_{bg}$ by our simple model is a consequence of our setting $\alpha = \alpha_1$ before differentiating the charge expressions, as discussed in Sec. 8.3. If we use instead $\alpha < \alpha_1$, as is commonly done in conjunction with drain current modeling, we find $C_{bg} < C_{gb}$ ($C_{mx} < 0$), i.e., a conclusion which is opposite from the result obtained by measurements and accurate models—a rather unexpected error.



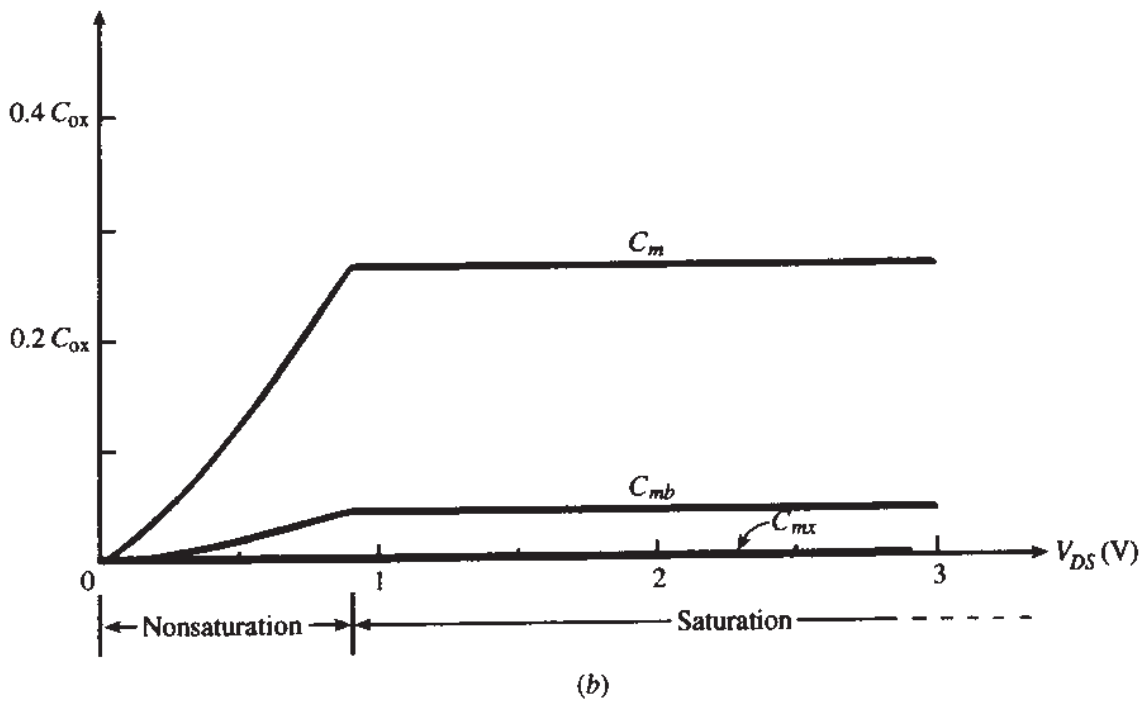
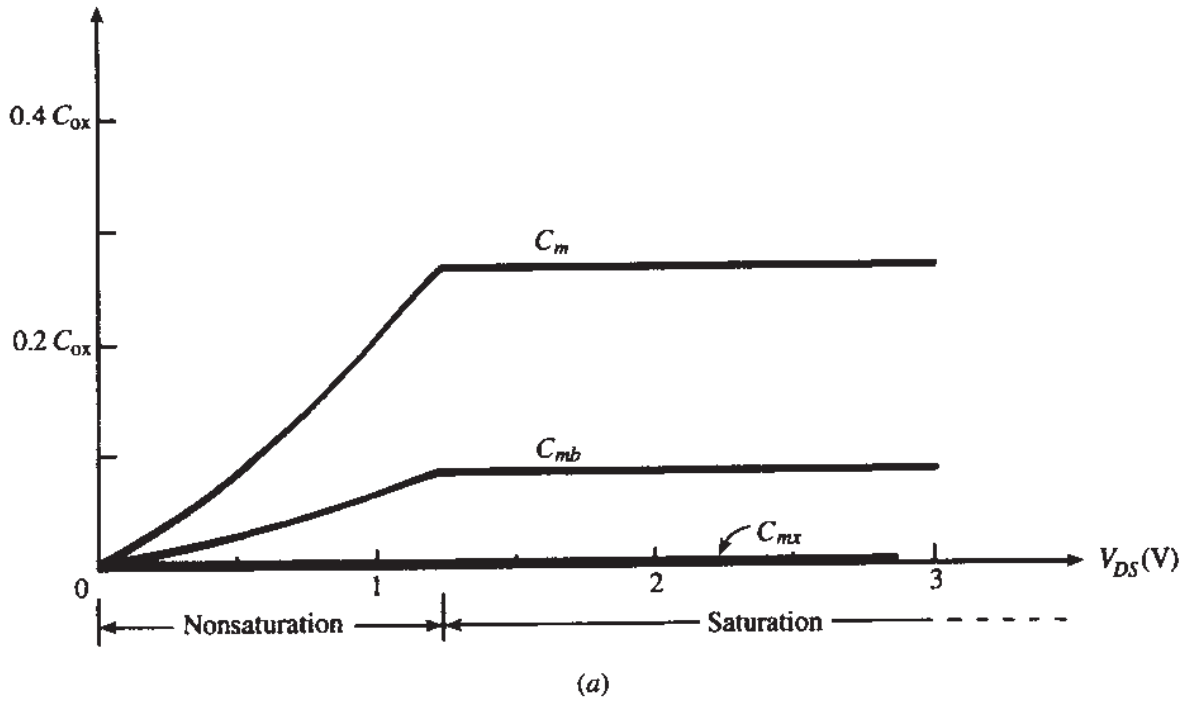
(a)



(b)

FIGURE 9.7

Drain-gate, drain-substrate, substrate-gate, and source-drain small-signal capacitances for a device with $V_{T0} = 0.5$ V, $\gamma = 0.6$ V^{1/2}, $\phi_0 = 0.9$ V, with $V_{GS} = 2$ V, as obtained from (9.2.21) to (9.2.24). (a) $V_{SB} = 0$; (b) $V_{SB} = 2$ V.

**FIGURE 9.8**

Small-signal transcapacitances, C_m , C_{mb} , and C_{mx} for a device with $V_{T0} = 0.5$ V, $\gamma = 0.6$ V^{1/2}, $\phi_0 = 0.9$ V, with $V_{GS} = 2$ V, as obtained from (9.2.25) to (9.2.27). (a) $V_{SB} = 0$; (b) $V_{SB} = 2$ V.

Using results from above and (8.3.15), one obtains (if V_{DS} or V_{GS} is small and/or V_{SB} is large)

$$\frac{C_{db}}{C_{dg}} \approx \frac{C_{sb}}{C_{sg}} \approx \frac{C_{bd}}{C_{gd}} \approx \frac{C_{bs}}{C_{gs}} \approx \frac{C_{bb}}{C_{gg}} \approx \frac{C_{mb}}{C_m} \approx \frac{C'_{bc}(V_{SB})}{C'_{ox}} \approx \frac{g_{mb}}{g_m} \approx \frac{dV_T}{dV_{SB}} = \alpha_1 - 1 \quad (9.2.30)$$

where we have used (8.3.13b).

We consider now the values of the capacitance parameters in two special cases of interest.

Nonsaturation with $V_{DS} = 0$: Using $V_{DS} = 0$ ($\eta = 1$) in the above results, we easily obtain the following values (below we also repeat the results obtained in Sec. 8.3 for completeness):

$$C_{gg} = C_{ox} = C'_{ox} WL \quad (9.2.31a)$$

$$C_{dg} = C_{gd} = C_{gs} = C_{sg} = \frac{C_{gg}}{2} \quad (9.2.31b)$$

$$C_{bb} = (\alpha_1 - 1) C_{gg} = C'_{bc}(V_{SB}) WL \quad (9.2.31c)$$

$$C_{db} = C_{bd} = C_{bs} = C_{sb} = \frac{C_{bb}}{2} \quad (9.2.31d)$$

$$C_{gb} = C_{bg} = 0 \quad (9.2.31e)$$

$$C_{dd} = C_{ss} = \alpha_1 \frac{C_{ox}}{3} \quad (9.2.31f)$$

$$C_{ds} = C_{sd} = -\alpha_1 \frac{C_{ox}}{6} \quad (9.2.31g)$$

$$C_m = C_{mb} = C_{mx} = 0 \quad (9.2.31h)$$

We note that, for the case we are considering ($V_{DS} = 0$), we have $C_{kl} = C_{lk}$, where k and l represent any two among s , d , g , and b . This is actually the only case where such an equality is observed exactly by all C_{kl} ; with $V_{DS} \neq 0$ one finds $C_{kl} \neq C_{lk}$ (for $k \neq l$) in general, for reasons already discussed.

Some of the above results have already been obtained and discussed in Sec. 8.3. Here we will attempt to make plausible the rest of them. The value of the inversion layer charge per unit area, Q'_l at the source and drain ends of the channel is, from Sec. 3.4.2,

$$Q'_{l0} = -C'_{ox}(V_{GS} - V_{TS}) \quad (9.2.32a)$$

$$Q'_{lL} = -C'_{ox}(V_{GD} - V_{TD}) \quad (9.2.32b)$$

where

$$V_{TS} = V_{FB} + \phi_0 + \gamma\sqrt{V_{SB} + \phi_0} \quad (9.2.33a)$$

$$V_{TD} = V_{FB} + \phi_0 + \gamma\sqrt{V_{DB} + \phi_0} \quad (9.2.33b)$$

Of course, here we are interested in the case $V_{DS} = 0$. However, we have allowed above for independent voltages V_D and V_S , so that we can differentiate the expressions involved in terms of one voltage while keeping the other constant in accordance with the capacitance definitions.

For $V_D \approx V_S$, one can assume a straight-line variation of Q'_l with the position x along the channel:

$$Q'_l = Q'_{l0} + (Q'_{lL} - Q'_{l0}) \frac{x}{L} \quad (9.2.34)$$

One can now find Q_D and Q_S by using the above in (7.3.9). The result is

$$Q_D = -C_{ox} \left[\frac{1}{6} (V_{GS} - V_{TS}) + \frac{1}{3} (V_{GD} - V_{TD}) \right] \quad (9.2.35)$$

$$Q_S = -C_{ox} \left[\frac{1}{3} (V_{GS} - V_{TS}) + \frac{1}{6} (V_{GD} - V_{TD}) \right] \quad (9.2.36)$$

Various capacitance values in (9.2.31) can now be verified easily by using the above expressions. In particular, the reason for the denominators 3 and 6 in (9.2.31f) and (9.2.31g) becomes apparent.

All results in (9.2.31) agree exactly with those obtained by using the charges corresponding to the complete strong-inversion model.

Saturation: Using the general capacitance expressions with $V_{DS} = V'_{DS}$ ($\eta = 0$) gives the following results (some of them have already been derived in Sec. 8.3 and are repeated below for convenience).

$$C_{dg} = \frac{4}{15} C_{ox} \quad (9.2.37a)$$

$$C_{gd} = 0 \quad (9.2.37b)$$

$$C_{db} = (\alpha_1 - 1)C_{dg} = \frac{4}{15} C'_{bc}(V_{SB})WL \quad (9.2.37c)$$

$$C_{bd} = 0 \quad (9.2.37d)$$

$$C_{sg} = \frac{2}{5} C_{ox} \quad (9.2.37e)$$

$$C_{gs} = \frac{2}{3} C_{ox} \quad (9.2.37f)$$

$$C_{sb} = (\alpha_1 - 1)C_{sg} = \frac{2}{5} C'_{bc}(V_{SB})WL \quad (9.2.37g)$$

$$C_{bs} = (\alpha_1 - 1)C_{gs} = \frac{2}{3} C'_{bc}(V_{SB})WL \quad (9.2.37h)$$

$$C_{gb} = C_{bg} = \frac{\alpha_1 - 1}{3\alpha_1} C_{ox} \quad (9.2.37i)$$

$$C_{ds} = -\alpha_1 \frac{4}{15} C_{ox} \quad (9.2.37j)$$

$$C_{sd} = 0 \quad (9.2.37k)$$

$$C_{gg} = \left[\frac{2}{3} + \frac{\alpha_1 - 1}{3\alpha_1} \right] C_{ox} \quad (9.2.37l)$$

$$C_{bb} = \left[\frac{2}{3}(\alpha_1 - 1) + \frac{\alpha_1 - 1}{3\alpha_1} \right] C_{ox} \quad (9.2.37m)$$

$$C_{dd} = 0 \quad (9.2.37n)$$

$$C_{ss} = \alpha_1 \frac{2}{5} C_{ox} \quad (9.2.37o)$$

$$C_m = \frac{4}{15} C_{ox} \quad (9.2.37p)$$

$$C_{mb} = (\alpha_1 - 1)C_m = \frac{4}{15} C'_{bc}(V_{SB})WL \quad (9.2.37q)$$

$$C_{mx} = 0 \quad (9.2.37r)$$

Note that in saturation we have $C_{kl} \neq C_{lk}$ (for $k \neq l$), which is to be expected, for the reasons discussed following (9.2.2). Note also that in saturation $C_{gd} = C_{bd} = C_{sd} = 0$, which is a manifestation of the fact that in this region the drain voltage does not influence the device charges.

WEAK INVERSION. The most important intrinsic capacitance in weak inversion, C_{gb} , has been discussed in Sec. 8.3. The other intrinsic capacitances in Fig. 9.5 are small, and their effect is usually swamped by that of extrinsic capacitances (Sec. 8.4) for all but very long devices.

GENERAL MODEL VALID IN ALL REGIONS OF INVERSION. Capacitances derived by using the accurate charge calculations suggested in Sec. 7.4.5 are shown in

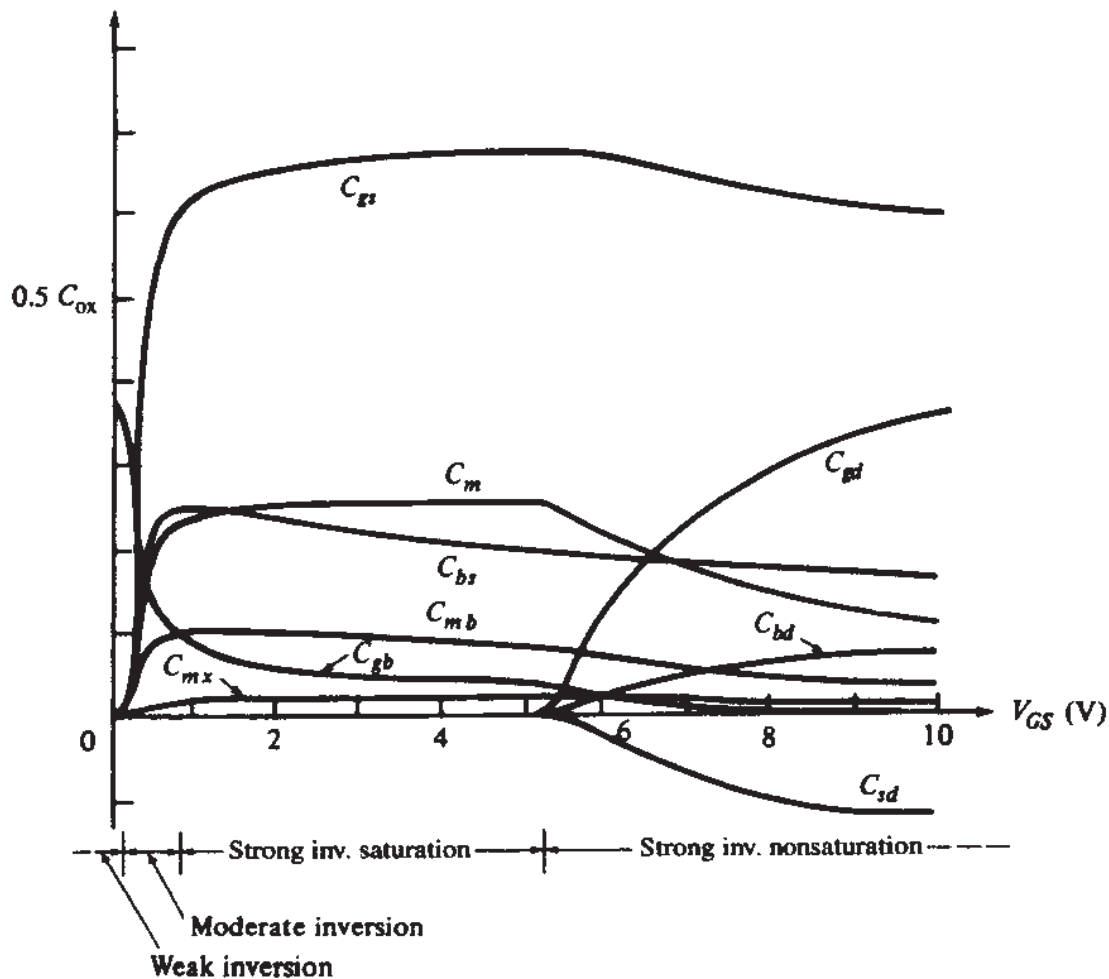


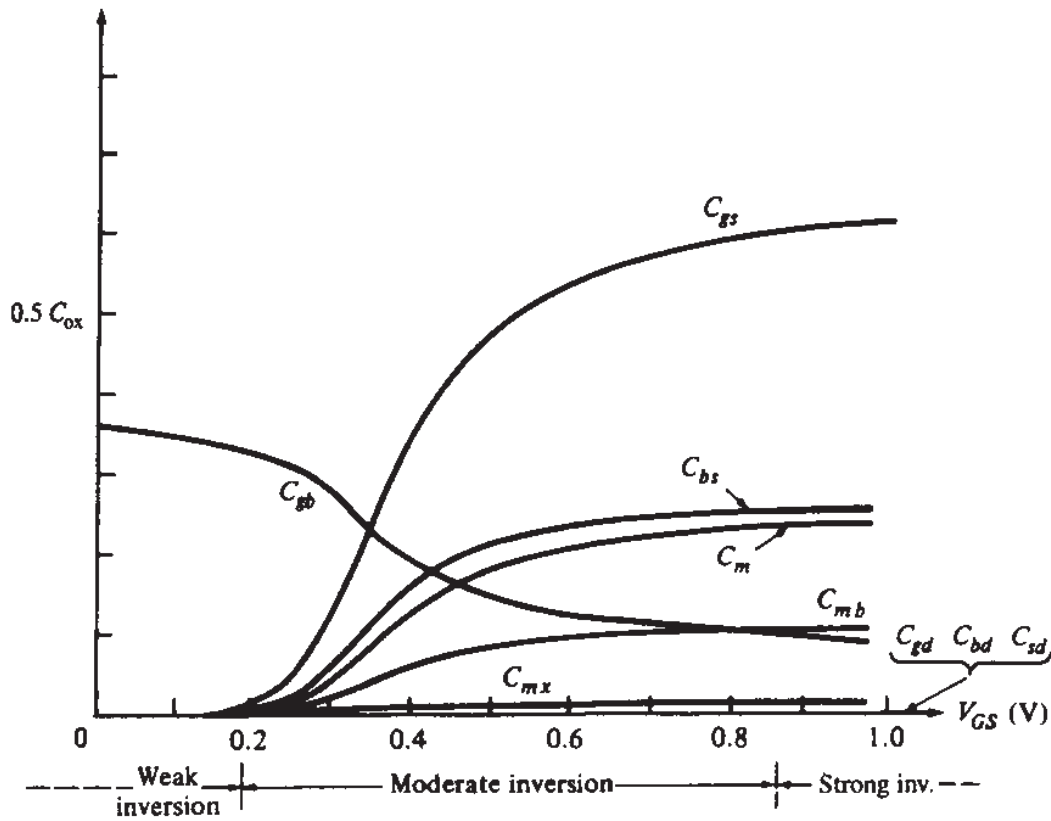
FIGURE 9.9

The nine capacitances in the small-signal model of Fig. 9.5, for the device of Fig. 8.21, plotted vs. V_{GS} for $V_{DS} = 4$ V and $V_{SB} = 0$, using accurate calculations.³⁶

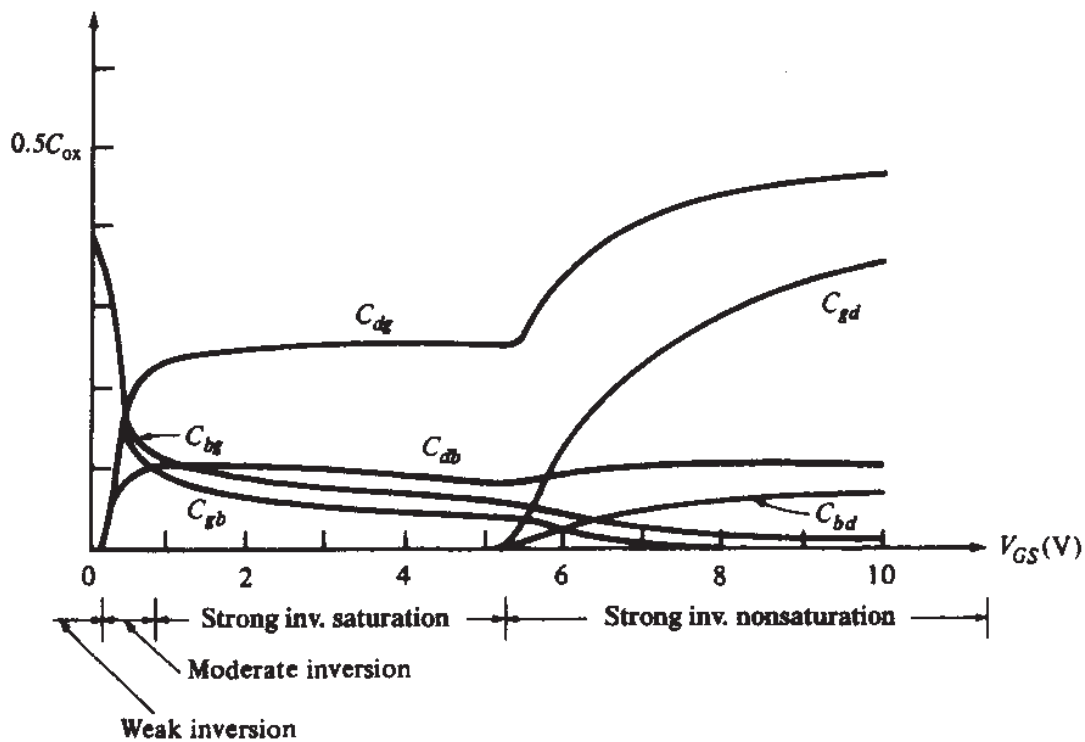
Fig. 9.9 versus V_{GS} .³⁶ All nine capacitances used in the model of Fig. 9.5 are shown. An expansion of the horizontal axis around the moderate-inversion region results in Fig. 9.10. Finally, Fig. 9.11 compares C_{gd} to C_{dg} , C_{bd} to C_{db} , and C_{gb} to C_{bg} . As seen, the capacitances in each of these pairs are, in general, different, as predicted by the theory in this section. These predictions agree with measurements.^{29,32,56} An example is shown in Fig. 9.12.

9.2.4 Frequency Region of Validity

The inclusion of the four capacitances C_m , C_{mb} , C_{mx} , and C_{sd} (Fig. 9.5) makes the complete quasi-static model better than the simple model of Fig. 8.17, as far as the frequency region of validity is concerned. The improvement depends on bias and on the terminals we are considering. For example, C_m is maximum in saturation, and it is there that the difference between the two models will be maximum, with respect to gate-to-drain action. At $V_{DS} = 0$, $C_m = 0$ and the two models are identical in that respect. On the other hand, it is at $V_{DS} = 0$ that C_{sd} is maximum, and hence it is at this point that the difference between the two models is maximum, as far as drain-to-source action is concerned.

**FIGURE 9.10**

The plot of Fig. 9.9 expanded around the moderate-inversion region.³⁶

**FIGURE 9.11**

A comparison of C_{dg} to C_{gd} , C_{db} to C_{bd} , and C_{bg} to C_{gb} for the device of Fig. 8.21, plotted versus V_{GS} for $V_{DS} = 4$ V and $V_{SB} = 0$, using accurate calculations.³⁶

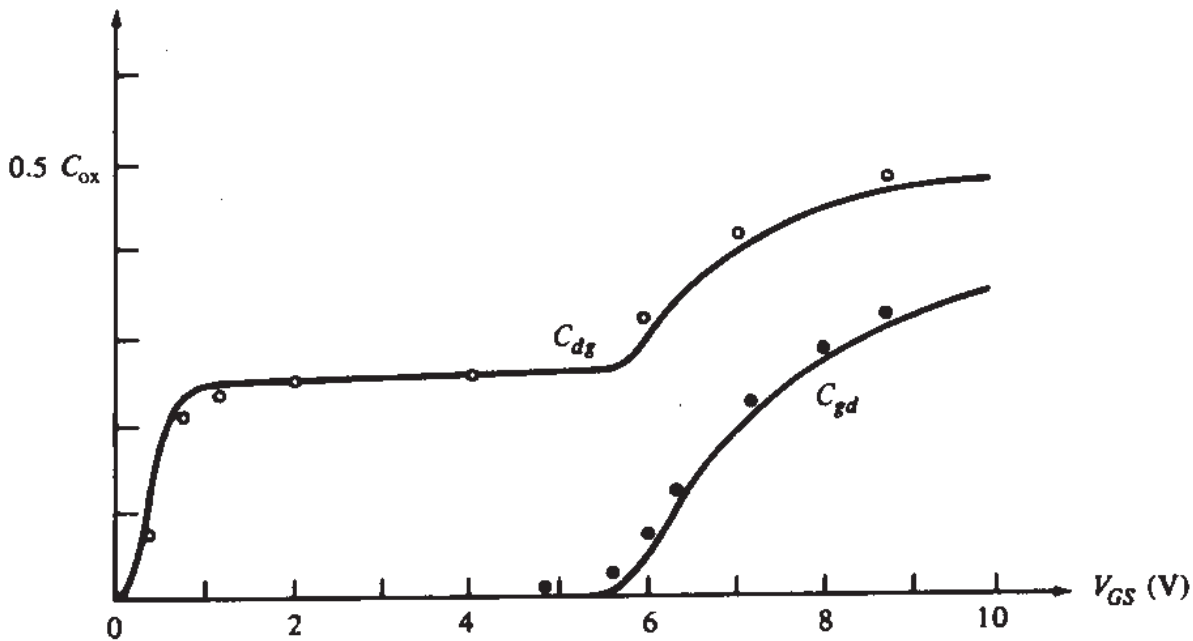


FIGURE 9.12

A comparison of C_{dg} to C_{gd} for the device of Fig. 8.21, plotted versus V_{GS} for $V_{DS} = 0$, and $V_{SB} = 0$ V. Points are measurements;³² lines are accurate calculations.³⁶

To provide some rough indication of the difference between the two models, consider operation in the saturation region, which is the most important region for small-signal circuit applications. Assume the element values in the complete quasi-static model of Fig. 9.5 have been chosen so that, at very low frequencies, the performance is good. Then the performance will continue to be good, with practically no deterioration up to about $\omega_o/3$, where ω_o is given by (8.3.6).³⁶ This result follows by comparison to more sophisticated, non-quasi-static models considered in Sec. 9.4. At this point, a warning should be given. *Although the frequency region of validity for the complete quasi-static model of Fig. 9.5 is larger, if this region is exceeded, this model can give very wrong results, in fact worse than those of the simple model of Fig. 8.17 in some respects.* This important point is considered in a model comparison in Sec. 9.4.4.

9.3 y-PARAMETER MODELS

Design of very-high-frequency circuits (not necessarily using MOS transistors) is often done by using the so-called “y parameters.” In this section we develop y-parameter models for the MOS device. We will first derive the general form of such models. In doing so, no assumptions will be made as to the physics of the device. In fact, we do not even have to assume the device is a MOS transistor. The only restriction we will place on it is that it has four terminals. For later convenience we will denote these terminals by D , G , B , and S , but for the present we do not have to associate these symbols with particular terminals of a specific device; indeed, we might as well have used X , Y , Z , and W instead.

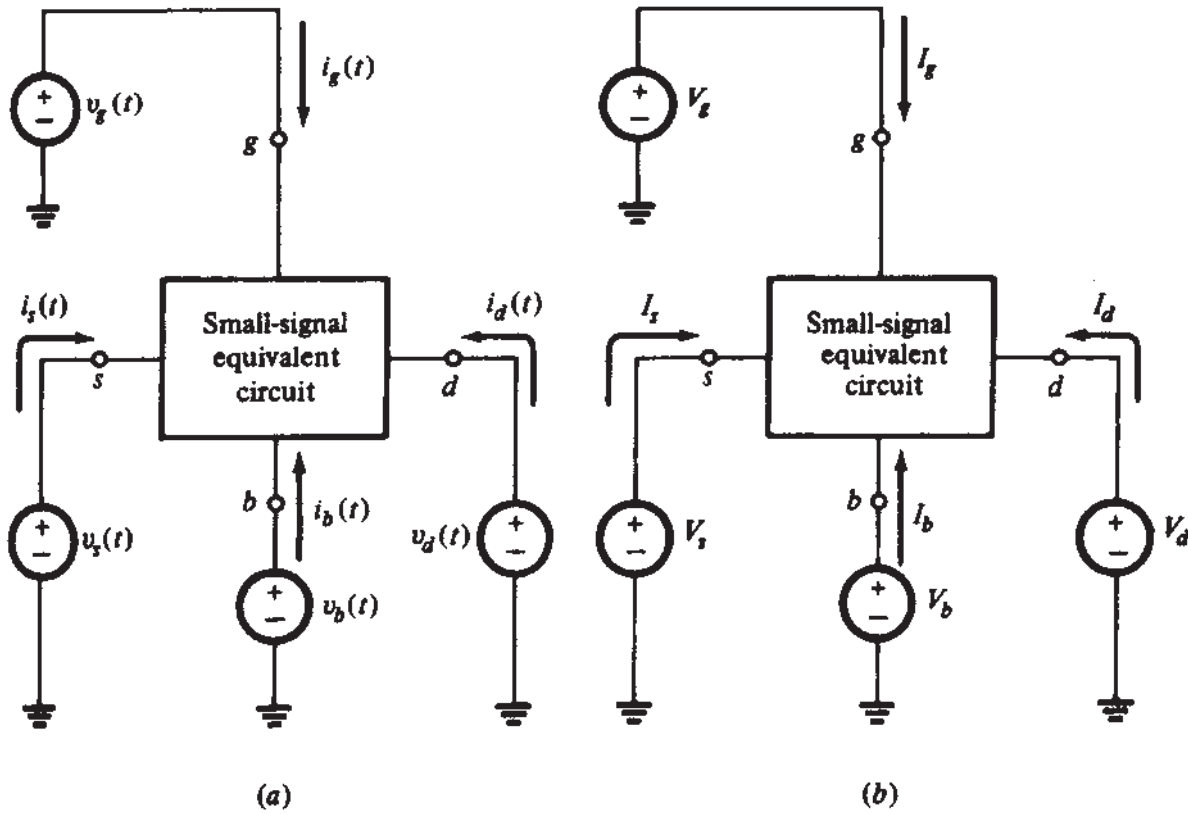


FIGURE 9.13

(a) The small-signal equivalent of Fig. 9.1b in the time domain; (b) corresponding representation in the frequency domain using phasors.

Let us consider the transistor driven by bias and small-signal voltages at each terminal, as shown in Fig. 9.1b. The small-signal equivalent circuit of the transistor driven by the small-signal parts of the voltage excitations is shown in Fig. 9.13a. We assume now that all small-signal voltages are sinusoidal and of the *same* angular frequency ω . Then in the *sinusoidal steady state* all small-signal currents will be sinusoidal and of the same frequency.⁵⁷⁻⁵⁹ The small-signal voltages and currents can be represented by cosine functions, e.g.,

$$v_g(t) = M_{vg} \cos(\omega t + \phi_{vg}) \quad (9.3.1)$$

We will use a *phasor* representation for each small-signal voltage or current, i.e., a complex number with magnitude and angle equal to the amplitude and phase, respectively, of the corresponding cosine waveform.⁵⁷⁻⁵⁹ Phasors will be denoted by capital letters with lowercase subscripts. For example, corresponding to $v_g(t)$ above we have a phasor V_g :

$$V_g = M_{vg} e^{j\phi_{vg}} \quad (9.3.2)$$

From now on we will use for brevity the terms “voltage” and “current” instead of the more complete terms “voltage phasor” and “current phasor.” Because the context will be clear, no confusion should arise. The phasor representation for the circuit of Fig. 9.13a is shown in Fig. 9.13b.

Let us assume that we are interested in the effect of V_g , V_b , V_d , and V_s on the current I_d . We will perform four experiments. In each, we will consider only one of the four small-signal voltages by setting the other three equal to zero in Fig. 9.13b. This is equivalent to setting to zero the values of three of the small-signal voltage sources in Fig. 9.1b, but, of course, *leaving all four dc bias sources intact* in that figure. The four experiments are summarized in Fig. 9.14. In each one, the ratio of the current phasor to the voltage phasor is a complex *admittance*.⁵⁷⁻⁵⁹ We will use the symbols shown in Fig. 9.14 for the four admittances.

Small-signal equivalent circuits are *linear* circuits, representing the fact that in an actual transistor with vanishingly small signals the quantities I_d , I_g , I_b , and I_s are *linear* functions of V_d , V_g , V_b , and V_s . One can thus use *superposition* to find I_d when all four of the small-signal voltages are active (nonzero). This is done by considering one voltage active at a time, evaluating the resulting current, and then adding up the currents:

$$I_d = I_d|_{V_g, V_b, V_s=0} + I_d|_{V_d, V_b, V_s=0} + I_d|_{V_d, V_g, V_s=0} + I_d|_{V_d, V_g, V_b=0} \quad (9.3.3)$$

Using the definitions in Fig. 9.14, the above equation can be written as follows:

$$I_d = y_{dd}V_d + y_{dg}V_g + y_{db}V_b + y_{ds}V_s \quad (9.3.4)$$

We can perform similar experiments to determine the currents in each of the other three terminals. If, in each case, we define admittances as follows:

$$y_{kl} = \frac{I_k}{V_l} \bigg|_{V_n=0, n \neq l} \quad (9.3.5)$$

we have a total of four equations:

$$I_d = y_{dd}V_d + y_{dg}V_g + y_{db}V_b + y_{ds}V_s \quad (9.3.6a)$$

$$I_g = y_{gd}V_d + y_{gg}V_g + y_{gb}V_b + y_{gs}V_s \quad (9.3.6b)$$

$$I_b = y_{bd}V_d + y_{bg}V_g + y_{bb}V_b + y_{bs}V_s \quad (9.3.6c)$$

$$I_s = y_{sd}V_d + y_{sg}V_g + y_{sb}V_b + y_{ss}V_s \quad (9.3.6d)$$

Note that these equations are similar in structure to (9.2.2), except that here no minus signs are used. This is a consequence of our definition (9.3.5) [which should be compared to (9.2.1)]. This definition of admittance parameters is standard in circuit theory. Equation (9.3.6), expressed in matrix form, is known as a *terminal*, or *indefinite, admittance matrix* representation.⁵⁹

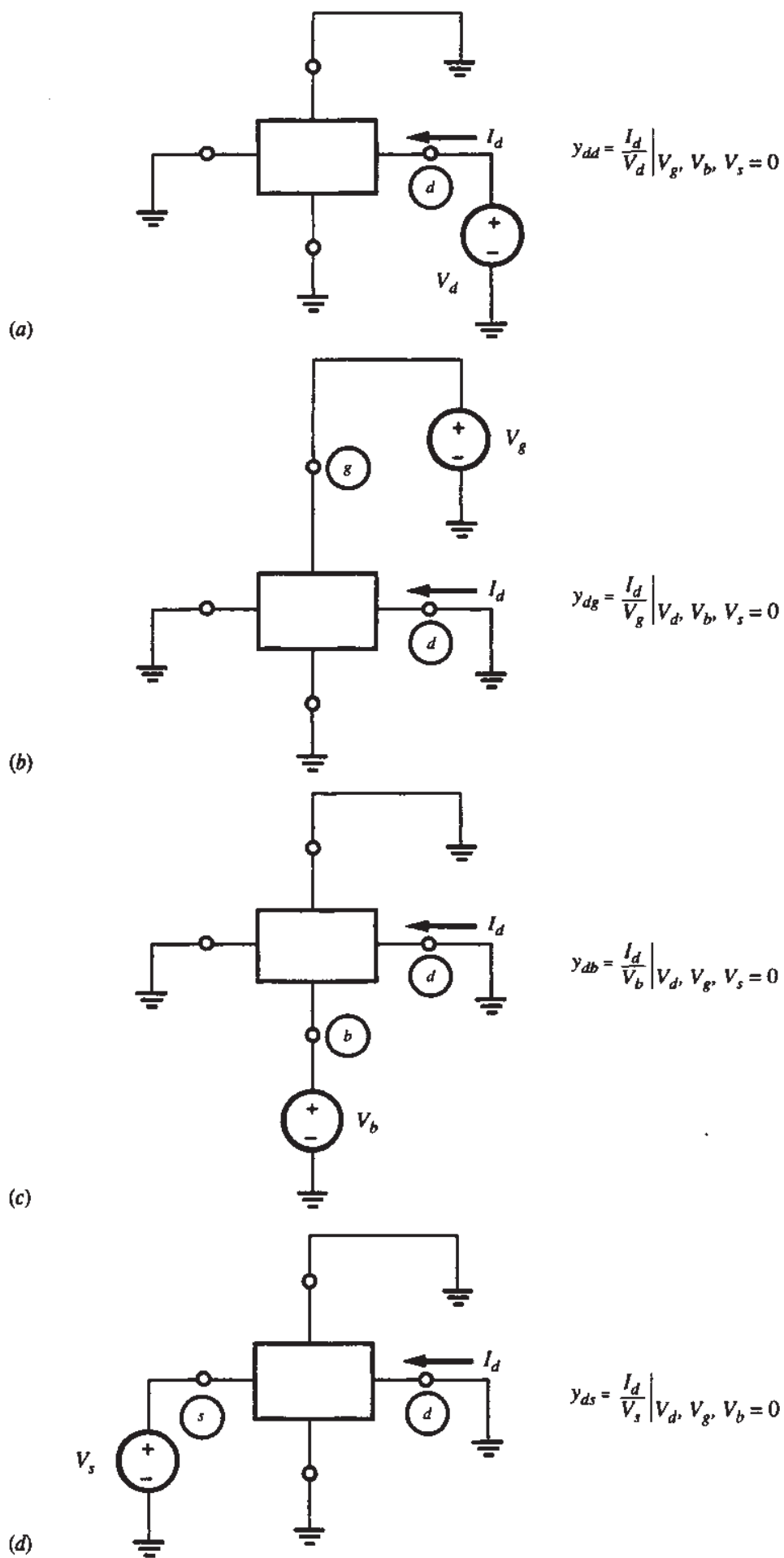


FIGURE 9.14

Definition of y parameters associated with the drain current. The notation $V_x, V_y, V_z = 0$ means $V_x = 0, V_y = 0, V_z = 0$.

We can follow a reasoning analogous to the one that led to (9.2.8) to get relations between the y parameters (Prob. 9.11):

$$y_{dd} + y_{dg} + y_{db} + y_{ds} = y_{dd} + y_{gd} + y_{bd} + y_{sd} = 0 \quad (9.3.7a)$$

$$y_{gg} + y_{gd} + y_{gb} + y_{gs} = y_{gg} + y_{dg} + y_{bg} + y_{sg} = 0 \quad (9.3.7b)$$

$$y_{bb} + y_{bd} + y_{bg} + y_{bs} = y_{bb} + y_{db} + y_{gb} + y_{sb} = 0 \quad (9.3.7c)$$

$$y_{ss} + y_{sd} + y_{sg} + y_{sb} = y_{ss} + y_{ds} + y_{gs} + y_{bs} = 0 \quad (9.3.7d)$$

Similarly, following a reasoning analogous to the one that led to (9.2.12), we conclude that the fourth equation in (9.3.6) can be omitted (in fact, any one among the four equations could have been chosen for omission) without losing information, and that the remaining three equations can be written as follows:

$$I_d = y_{dd}V_{ds} + y_{dg}V_{gs} + y_{db}V_{bs} \quad (9.3.8a)$$

$$I_g = y_{gd}V_{ds} + y_{gg}V_{gs} + y_{gb}V_{bs} \quad (9.3.8b)$$

$$I_b = y_{bd}V_{ds} + y_{bg}V_{gs} + y_{bb}V_{bs} \quad (9.3.8c)$$

with $V_{kl} = V_k - V_l$. The above set of equations can be represented by the circuit of Fig. 9.15, a fact verifiable directly by writing Kirchhoff's law at nodes d , g , and b for the current phasors.

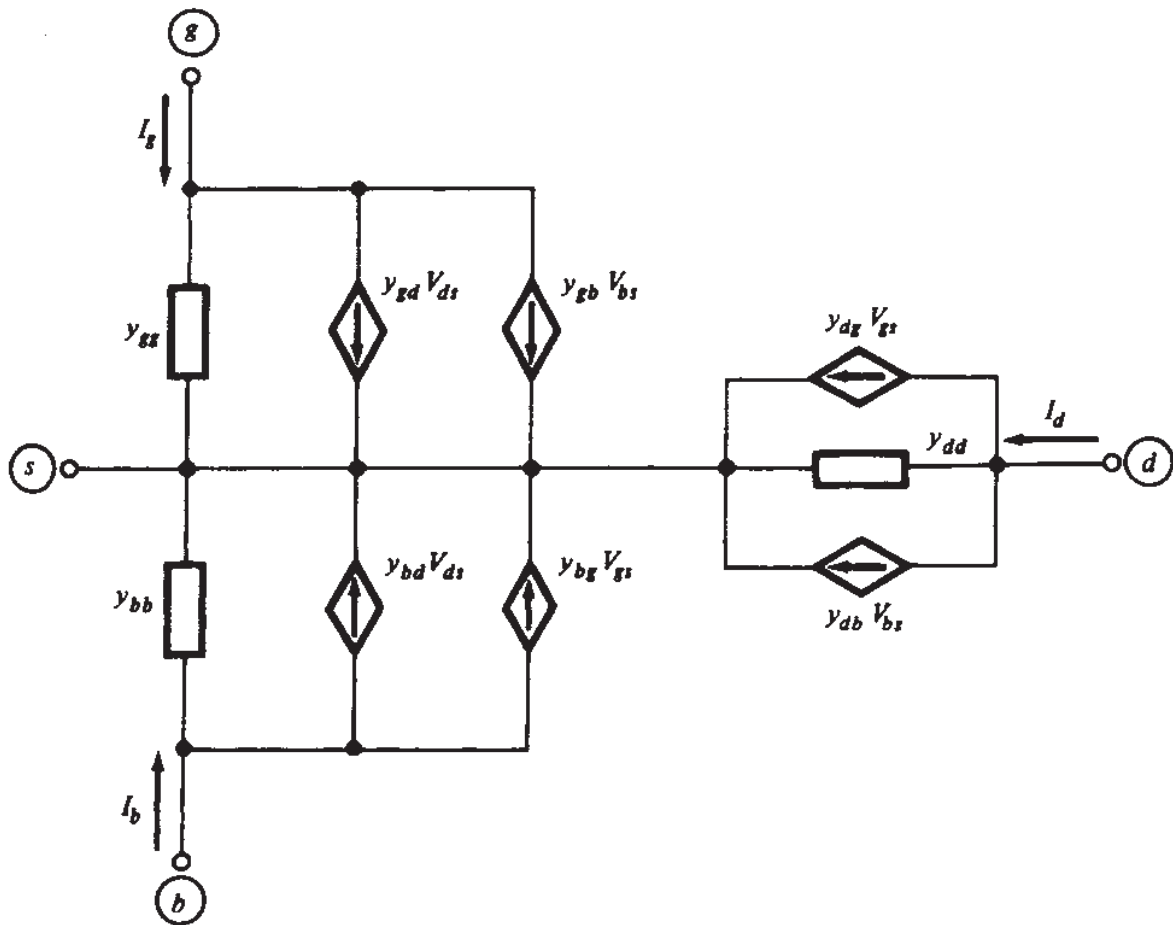
Other three-port y -parameter representations are also possible. For example, if instead of using the s terminal as a potential reference and omitting (9.3.6d) we had used the b terminal as a reference and had omitted (9.3.6c), we would have obtained the representation shown in Fig. 9.16. This representation has an appealing symmetry. For a symmetrically laid out device, the role of source and drain is identical; thus $y_{ss} = y_{dd}$, $y_{sg} = y_{dg}$, and $y_{sd} = y_{ds}$. Using such equal values in Fig. 9.16 makes the symmetry evident. Nevertheless, small-signal models using the substrate as the reference are not in much use for high-frequency work. The relative merits of using the source or the substrate as a reference are discussed in Sec. 4.9.

We now go one step closer in relating our present models to the one in Fig. 9.5. Following an approach similar to the one that led to (9.2.19), we can rewrite (9.3.8) as follows (Prob. 9.12):

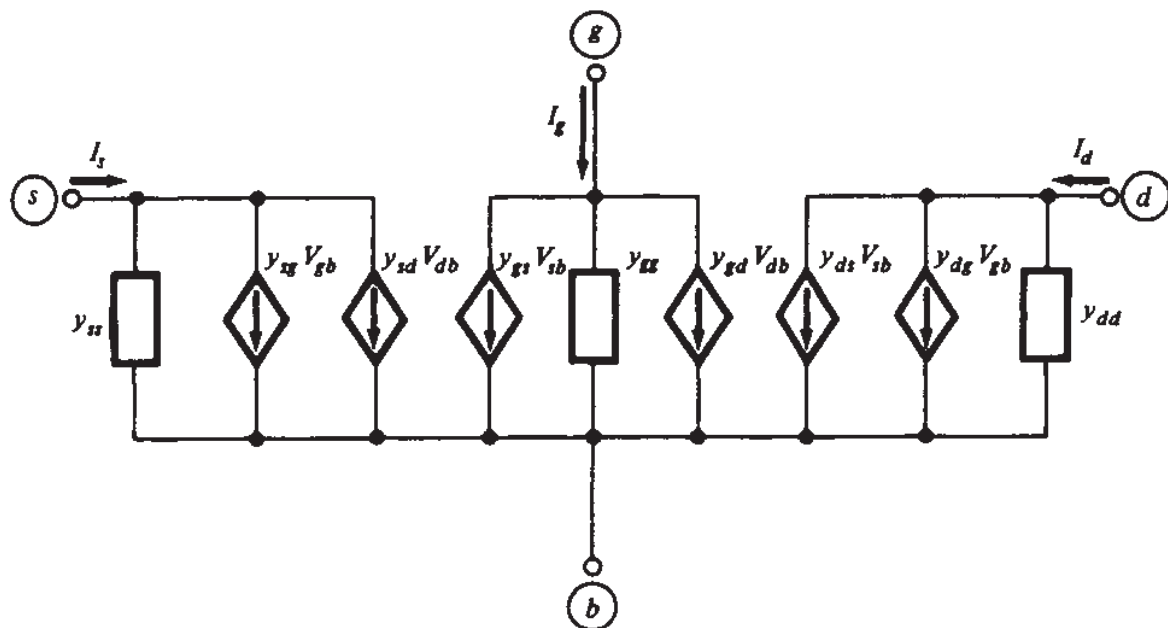
$$I_d = -y_{gd}V_{dg} - y_{sd}V_{ds} - y_{bd}V_{db} + y_mV_{gs} + y_{mb}V_{bs} \quad (9.3.9a)$$

$$I_g = -y_{gd}V_{gd} - y_{gb}V_{gb} - y_{gs}V_{gs} \quad (9.3.9b)$$

$$I_b = -y_{bd}V_{bd} - y_{gb}V_{bg} + y_{mx}V_{gb} - y_{bs}V_{bs} \quad (9.3.9c)$$


FIGURE 9.15

A general y -parameter model using the source terminal as a reference.


FIGURE 9.16

A general y -parameter model using the substrate terminal as a reference.

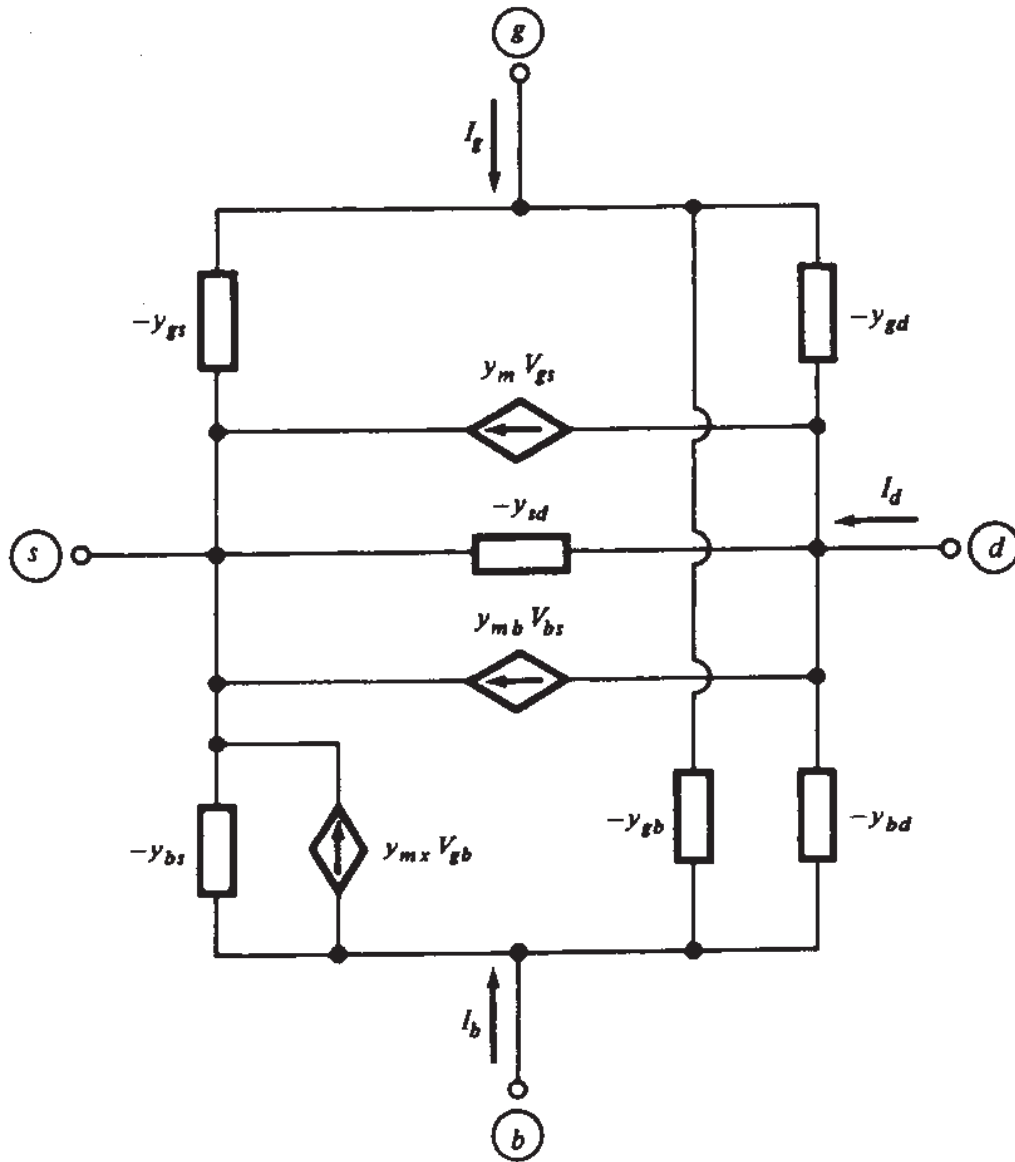


FIGURE 9.17
A general y -parameter model.³⁹

where

$$y_m = y_{dg} - y_{gd} \quad (9.3.10a)$$

$$y_{mb} = y_{db} - y_{bd} \quad (9.3.10b)$$

$$y_{mx} = y_{bg} - y_{gb} \quad (9.3.10c)$$

These equations can be represented by the circuit of Fig. 9.17, as can be verified by writing Kirchhoff's current law for nodes d , g , and b . Since the development of this model was general, the model in Fig. 9.5 should simply be a special case of it. By comparing the two circuits we get,[†] for this special case,

[†]To arrive at (9.3.11) we use the following fact:⁵⁷⁻⁵⁹ The time-domain i - v equation for a capacitor, namely $i(t) = C dv(t)/dt$, corresponds to the phasor equation $I = j\omega CV$, where $j\omega C$ is the admittance corresponding to the capacitance C .

$$-y_{gd} = j\omega C_{gd} \quad (9.3.11a)$$

$$-y_{gs} = j\omega C_{gs} \quad (9.3.11b)$$

$$-y_{bd} = j\omega C_{bd} \quad (9.3.11c)$$

$$-y_{bs} = j\omega C_{bs} \quad (9.3.11d)$$

$$-y_{gb} = j\omega C_{gb} \quad (9.3.11e)$$

$$-y_{sd} = g_{sd} + j\omega C_{sd} \quad (9.3.11f)$$

$$y_m = g_m - j\omega C_m \quad (9.3.11g)$$

$$y_{mb} = g_{mb} - j\omega C_{mb} \quad (9.3.11h)$$

$$y_{mx} = -j\omega C_{mx} \quad (9.3.11i)$$

In the above equations, observe that: (1) y_m , y_{mb} , and $-y_{sd}$ have a constant, real, positive part, and (2) all y parameters have imaginary parts which are proportional to frequency. These two observations are, indeed, verified by measurements at frequencies up to about $\omega_o/3$, with ω_o as given by (8.3.6). Indeed, the capacitance values can be determined as the constant of proportionality in the imaginary part of measured admittances.^{32,38} At frequencies higher than $\omega_o/3$, however, the behavior predicted above fails. Measurements then show, for example, that both the real and the imaginary part of y_m decrease in magnitude with frequency, and that y_{gs} begins to have a nonzero real part. To explain such phenomena, the quasi-static assumption must be abandoned. This is what will be done in the following section.

9.4 NON-QUASI-STATIC MODELS

9.4.1 Introduction

In this section, the quasi-static assumption will be dropped and the dynamics of the channel charge at high frequencies will be investigated in detail. Before starting a mathematical analysis, we present a short discussion as to what one might expect intuitively at frequencies where quasi-static behavior is no longer observed.

Let us consider the intrinsic transistor in Fig. 9.18. We assume that, in general, each terminal voltage consists of a dc bias and a small-signal sinusoidal component, as shown. Assume that only one small-signal voltage is nonzero at a time. If v_s is varying very slowly, the inversion layer charge has time to follow with practically no delay. We have seen that the resulting effect on the gate can be modeled by connecting a capacitance C_{gs} from source to gate. However, if the variation of v_s is fast, the “inertia” of the inversion layer becomes nonnegligible, and the effect (gate charge change) will lag behind the cause (source voltage change). A similar effect will be observed between drain and gate. Also, similar conclusions can be drawn for the effect of the source and of the drain on the substrate (the “back gate”). Consider now the effect of the gate voltage. If v_g is varying very fast, the inversion layer charge does not have enough time to respond fully, and thus $|y_{dg}|$, which models this response, will be small. Also, the angle of y_{dg} should be significant and negative, because of the delay

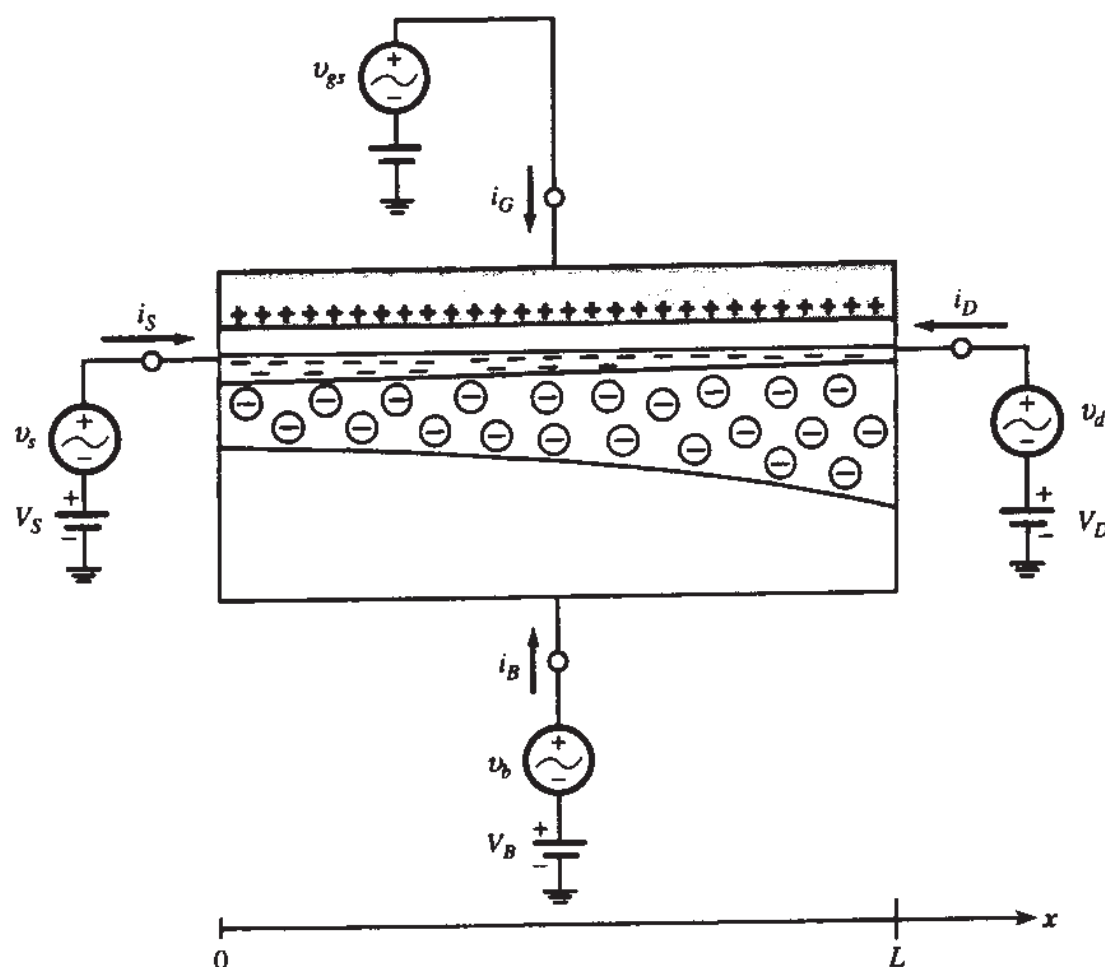


FIGURE 9.18
Intrinsic transistor with bias and small-signal voltages.

between the cause (the variation in the gate voltage) and the effect (the variation in the drain current). Finally, similar observations hold for the effect of the substrate voltage on the inversion layer charge. Measurements⁵¹ on both long- and short-channel devices do, in fact, verify that such effects can influence transistor behavior.

The above effects will be observed if the operating frequency exceeds the upper limit of validity for quasi-static modeling. From Sec. 9.2.4, this upper limit is proportional to ω_o , which in turn is proportional to $1/L^2$ in the absence of velocity saturation. One way to model the transistor at frequencies above the limit is, in principle, to split the device into sections (Fig. 7.12). The length of each section is chosen such that, for it, a quasi-static model *can* be used. The combination of the models of all sections will then be a valid model for the whole transistor at the frequency of interest. The higher this frequency, the shorter the length needed for each section. One is thus led to carrying the idea to the limit, i.e., considering elemental sections the lengths of which are allowed to approach zero. The following analysis corresponds mathematically to this idea. The analysis is based on well-established principles,^{1,2,4,17,20,21,35} but is generalized to view the transistor as a four-terminal device.³⁹ We want to establish all steps carefully, so the procedure will be rather long; once again we ask for the reader's patience.

9.4.2 A Non-Quasi-Static Strong-Inversion Model

In this section we will derive a useful high-frequency model corresponding to the simplified strong-inversion dc model. To end up with manageable expressions, we will use here the same simplifying assumptions we used in Sec. 8.3.³⁹ The first assumption is that, in the expressions for the charge, the quantity α will be assumed to be given by

$$\alpha = \alpha_1 = 1 + \frac{\gamma}{2\sqrt{\phi_0 + V_{SB}}} \quad (9.4.1)$$

The second assumption is that the derivative of α_1 with V_S or V_B is negligible for our purposes, and thus, α_1 will be treated as a constant during differentiation. At a later point, we will verify that the above assumptions do indeed produce a useful model by comparison to more exact models.

Many of the expressions we have developed in previous chapters will be needed here. For convenience, we will repeat these below. We will take this opportunity and present the various expressions in an organized fashion, so as to present a complete picture of, and the relation between, the following cases:

1. DC (bias) excitation
2. Time-varying excitation
3. The special case of (2) where the time variations are small signals
4. A special case of (3), where the small signals have a form particularly useful for high-frequency model development.

DC (BIAS) EXCITATION. In our analysis we will find it convenient to express the gate, depletion region, and inversion layer charges per unit area in terms of $V_{GS} = V_{GB} - V_{SB}$ and $V_{CS}(x) = V_{CB}(x) - V_{SB}$. Recall that, in strong inversion, $V_{CB}(x)$ can be viewed as the effective reverse bias between the inversion layer at point x and the substrate. Thus, $V_{CS}(x)$ represents the potential drop across the part of the inversion layer contained between point x and the source region. Using the above in (4.5.29) and (4.5.30), using (9.4.1), employing the charge conservation equation $Q'_G + Q'_o + Q'_I + Q'_B = 0$, and recalling (7.2.3), we have the following results.

The gate charge per unit area and the total gate charge are, respectively,

$$Q'_G(x) = C'_{ox}[V_{GS} - V_{FB} - \phi_0 - V_{CS}(x)] - Q'_o \quad (9.4.2)$$

$$Q_G = W \int_0^L Q'_G(x) dx \quad (9.4.3)$$

The corresponding charges for the depletion region are

$$Q'_B(x) = -C'_{ox}[\gamma\sqrt{\phi_0 - V_{BS}} + (\alpha_1 - 1)V_{CS}(x)] \quad (9.4.4)$$

$$Q_B = W \int_0^L Q'_B(x) dx \quad (9.4.5)$$

The inversion layer charge per unit area can be written in the form

$$Q'_I(x) = -C'_{ox} U_I(x) \quad (9.4.6)$$

where

$$U_I(x) = V_{GS} - V_{FB} - \phi_0 - \gamma\sqrt{\phi_0 - V_{BS}} - \alpha_1 V_{CS}(x) \quad (9.4.7)$$

The current at point x in the channel will be denoted by $I_I(x)$. This current is given by the right-hand side of (4.5.6). Using $V_{CS}(x) = V_{CB}(x) - V_{SB}$, we have

$$I_I(x) = -\mu W Q'_I(x) \frac{dV_{CS}(x)}{dx} \quad (9.4.8)$$

which, from (9.4.7), can be written as

$$I_I(x) = \frac{1}{\alpha_1} \mu W Q'_I(x) \frac{dU_I(x)}{dx} \quad (9.4.9)$$

and, since at dc the current is the same everywhere along the channel,

$$I_D = I_I(x) \quad (9.4.10)$$

Using (9.4.10) and (9.4.6) in (9.4.9), and integrating both sides from x to L , we get

$$I_D = \frac{W}{L-x} \frac{\mu C'_{ox}}{2\alpha_1} [U_I^2(x) - U_I^2(L)] \quad (9.4.11)$$

which for $x = 0$ gives

$$I_D = \frac{W}{L} \frac{\mu C'_{ox}}{2\alpha_1} [U_I^2(0) - U_I^2(L)] \quad (9.4.12)$$

From the above two equations, we can solve for $U_I(x)$:

$$U_I(x) = \left\{ U_I^2(0) + \frac{x}{L} [U_I^2(L) - U_I^2(0)] \right\}^{1/2} \quad (9.4.13)$$

At the source end of the channel we have

$$V_{CS}(0) = 0 \quad (9.4.14a)$$

thus

$$U_I(0) = V_{GS} - V_{FB} - \phi_0 - \gamma\sqrt{\phi_0 - V_{BS}} \quad (9.4.14b)$$

At the drain end of the channel we have

$$\begin{aligned} V_{CS}(L) &= V_{DS}, \quad V_{DS} \leq V'_{DS} \\ &= V'_{DS}, \quad V_{DS} > V'_{DS} \end{aligned} \quad (9.4.15a)$$

thus,

$$U_I(L) = V_{GS} - V_{FB} - \phi_0 - \gamma\sqrt{\phi_0 - V_{BS}} - \alpha_1 V_{DS}, \quad V_{DS} \leq V'_{DS} \quad (9.4.15b)$$

$$= V_{GS} - V_{FB} - \phi_0 - \gamma\sqrt{\phi_0 - V_{BS}} - \alpha_1 V'_{DS}, \quad V_{DS} > V'_{DS} \quad (9.4.15c)$$

Using the above relations for $U_I(0)$ and $U_I(L)$, it is easy to verify that (9.4.12) is identical to the simplified-model equation (4.5.37) (with $\alpha = \alpha_1$; this restriction will be removed later). Similarly, (9.4.13) is equivalent to the equation for the potential distribution corresponding to the simplified strong-inversion model, given by (4.5.49). The form we use for these equations here will be found convenient later in this section.

As usual, for the gate and substrate currents under dc excitation we assume

$$I_G = 0 \quad (9.4.16)$$

$$I_B = 0 \quad (9.4.17)$$

TIME-VARYING EXCITATION. We will denote total (large-signal) quantities by lower-case symbols with capital subscripts. We show below how equations (9.4.2) to (9.4.7) have to be modified for the case of time-varying voltages (time dependence is explicitly indicated throughout):

$$q'_G(x, t) = C'_{ox}[v_{GS}(t) - V_{FB} - \phi_0 - v_{CS}(x, t)] - Q'_o \quad (9.4.18)$$

where Q'_o is assumed fixed.

$$q_G(t) = W \int_0^L q'_G(x, t) dx \quad (9.4.19)$$

$$q'_B(x, t) = -C'_{ox}[\gamma\sqrt{\phi_0 - v_{BS}(t)} + (\alpha_1 - 1)v_{CS}(x, t)] \quad (9.4.20)$$

$$q_B(t) = W \int_0^L q'_B(x, t) dx \quad (9.4.21)$$

$$q'_I(x, t) = -C'_{ox}u_I(x, t) \quad (9.4.22)$$

where

$$u_I(x, t) = v_{GS}(t) - V_{FB} - \phi_0 - \gamma\sqrt{\phi_0 - v_{BS}(t)} - \alpha_1 v_{CS}(x, t) \quad (9.4.23)$$

Equation (9.4.9) must be replaced by

$$i_I(x, t) = \frac{\mu W}{\alpha_1} q'_I(x, t) \frac{\partial u_I(x, t)}{\partial x} \quad (9.4.24)$$

Since we will allow fast variations, (9.4.10) does not have a corresponding equation in the present case. Instead, we must consider the "continuity equation" (7.7.5). That equation, from (9.4.22), can be written as

$$\frac{\partial i_I(x, t)}{\partial x} = -C'_{ox} W \frac{\partial u_I(x, t)}{\partial t} \quad (9.4.25)$$

Note that for the special case of no variation with time, this equation is consistent with (9.4.10).

The terminal currents now are given by

$$i_D(t) = i_I(L, t) \quad (9.4.26)$$

$$i_G(t) = \frac{dq_G(t)}{dt} \quad (9.4.27)$$

$$i_B(t) = \frac{dq_B(t)}{dt} \quad (9.4.28)$$

SMALL-SIGNAL EXCITATION. We now assume that the total terminal voltages are of the following form:

$$v_{GS}(t) = V_{GS} + v_{gs}(t) \quad (9.4.29a)$$

$$v_{BS}(t) = V_{BS} + v_{bs}(t) \quad (9.4.29b)$$

$$v_{DS}(t) = V_{DS} + v_{ds}(t) \quad (9.4.29c)$$

where in each right-hand side the first term is a bias quantity, and the second term is a small signal (all small-increment quantities will be represented by lowercase subscripts). As a result of the above form of the voltages, we will have for the other time-varying quantities in our analysis:

$$q'_G(x, t) = Q'_G(x) + q'_g(x, t) \quad (9.4.30a)$$

$$q_G(t) = Q_G + q_g(t) \quad (9.4.30b)$$

$$q'_B(x, t) = Q'_B(x) + q'_b(x, t) \quad (9.4.30c)$$

$$q_B(t) = Q_B + q_b(t) \quad (9.4.30d)$$

$$u_I(x, t) = U_I(x) + u_i(x, t) \quad (9.4.30e)$$

$$v_{CS}(x, t) = V_{CS}(x) + v_{cs}(x, t) \quad (9.4.30f)$$

$$i_I(x, t) = I_I(x) + i_i(x, t) \quad (9.4.30g)$$

$$i_D(t) = I_D + i_d(t) \quad (9.4.30h)$$

$$i_G(t) = I_G + i_g(t) \quad (9.4.30i)$$

$$i_B(t) = I_B + i_b(t) \quad (9.4.30j)$$

Assuming $v_{gs}(t)$, $v_{bs}(t)$, and $v_{ds}(t)$ are very small, all quantities denoted above with capital letters and capital subscripts will be the same as under dc excitation; values for them have already been found. Using quantities from (9.4.29) and (9.4.30) in (9.4.18) to (9.4.28), one can split the resulting expressions into a "bias" and a small-increment part. For example, using (9.4.29a), (9.4.30a), and (9.4.30f) in (9.4.18), we have

$$\begin{aligned} Q'_G(x) + q'_g(x, t) &= C'_{ox}[V_{GS} + v_{gs}(t) - V_{FB} - \phi_0 - V_{CS}(x) - v_{cs}(x, t)] - Q'_o \\ &= \{C'_{ox}[V_{GS} - V_{FB} - \phi_0 - V_{CS}(x)] - Q'_o\} \\ &\quad + C'_{ox}[v_{gs}(t) - v_{cs}(x, t)] \end{aligned} \quad (9.4.31)$$

The quantity in { } can be recognized as $Q'_G(x)$ from (9.4.2). Hence (9.4.31) gives

$$q'_g(x, t) = C'_{ox}[v_{gs}(t) - v_{cs}(x, t)] \quad (9.4.32)$$

Similarly, using (9.4.30b) in (9.4.19), gives

$$\begin{aligned} Q_G + q_g(t) &= W \int_0^L [Q'_G(x) + q'_g(x, t)] dx \\ &= W \int_0^L Q'_G(x) dx + W \int_0^L q'_g(x, t) dx \end{aligned} \quad (9.4.33)$$

which, using (9.4.3), gives

$$q_g(t) = W \int_0^L q'_g(x, t) dx \quad (9.4.34)$$

To derive similar expressions for the depletion region charges, we first note that the term containing the square root in (9.4.20) becomes, from (9.4.29b),

$$\gamma \sqrt{\phi_0 - v_{BS}(t)} = \gamma \sqrt{(\phi_0 - V_{BS}) - v_{bs}(t)} \quad (9.4.35)$$

Since $v_{bs}(t)$ is small, the right-hand side can be approximated by the first two terms of a series expansion. This gives

$$\gamma \sqrt{\phi_0 - v_{BS}(t)} = \gamma \sqrt{\phi_0 - V_{BS}} - (\alpha_1 - 1)v_{bs}(t) \quad (9.4.36)$$

with α_1 as given by (9.4.1). Using the above in (9.4.20) and proceeding as before, we obtain, for the small-increment part of $q'_B(t)$:

$$q'_b(x, t) = (\alpha_1 - 1) C'_{ox}[v_{bs}(t) - v_{cs}(x, t)] \quad (9.4.37)$$

Following the approach that led to (9.4.34), for $q_b(t)$ we get

$$q_b(t) = W \int_0^L q'_b(x, t) dx \quad (9.4.38)$$

For $u_i(x, t)$, starting from (9.4.23), using (9.4.36), and proceeding as before, we easily get

$$u_i(x, t) = v_{gs}(t) + (\alpha_1 - 1) v_{bs}(t) - \alpha_1 v_{cs}(x, t) \quad (9.4.39a)$$

$$= [v_{gs}(t) - v_{cs}(x, t)] + (\alpha_1 - 1)[v_{bs}(t) - v_{cs}(x, t)] \quad (9.4.39b)$$

Since $v_{cs}(x, t)$ is zero at the source end and equal to $v_{ds}(t)$ at the drain end, we have, from the above equation,

$$u_i(0, t) = v_{gs}(t) + (\alpha_1 - 1)v_{bs}(t) \quad (9.4.40)$$

$$u_i(L, t) = [v_{gs}(t) - v_{ds}(t)] + (\alpha_1 - 1)[v_{bs}(t) - v_{ds}(t)] \quad (9.4.41)$$

The small-signal quantity $i_i(x, t)$ can be obtained by starting from (9.4.24). Using the fact that $u_i(x, t)$ is very small, the result can be put in the form (Prob. 9.13)

$$i_i(x, t) = -\frac{\mu W C'_{ox}}{\alpha_1} \frac{\partial}{\partial x} [U_I(x) u_i(x, t)] \quad (9.4.42)$$

Starting from (9.4.25) we get, using the facts that $\partial I_I(x)/\partial x = 0$ and $\partial U_I(x)/\partial t = 0$,

$$\frac{\partial i_i(x, t)}{\partial x} = -C'_{ox} W \frac{\partial u_i(x, t)}{\partial t} \quad (9.4.43)$$

For the drain small-signal current we have, starting from (9.4.26),

$$i_d(t) = i_i(L, t) \quad (9.4.44)$$

For the gate small-signal current, starting from (9.4.27), we obtain

$$i_g(t) = \frac{dq_g(t)}{dt} \quad (9.4.45)$$

We can now use (9.4.34) and (9.4.32) in this equation. If in the result we substitute $v_{cs}(x, t)$ as it ensues from solving (9.4.39), we obtain

$$i_g(t) = W C'_{ox} \frac{d}{dt} \int_0^L \left\{ \frac{\alpha_1 - 1}{\alpha_1} [v_{gs}(t) - v_{bs}(t)] + \frac{1}{\alpha_1} u_i(x, t) \right\} dx \quad (9.4.46)$$

For the substrate small-signal current we have, starting from (9.4.28),

$$i_b(t) = \frac{dq_b(t)}{dt} \quad (9.4.47)$$

Using in this equation (9.4.38) and (9.4.37), and substituting $v_{cs}(x, t)$ from (9.4.39) we obtain

$$i_b(t) = (\alpha_1 - 1)WC'_{ox} \frac{d}{dt} \int_0^L \left\{ \frac{1}{\alpha_1} [v_{bs}(t) - v_{gs}(t)] + \frac{1}{\alpha_1} u_i(x, t) \right\} dx \quad (9.4.48)$$

It is clear from the above expression that to find any of the terminal currents one needs an expression for $u_i(x, t)$, and that must be found from (9.4.42) and (9.4.43). The result, of course, depends on the form of the terminal small-signal voltages through the boundary conditions (9.4.40) and (9.4.41). In what follows, we will present the solution when the terminal voltages assume a form of special interest.

COMPLEX EXPONENTIAL EXCITATION. One could assume that the small-signal voltages are sinusoids and consider the corresponding small-signal terminal currents in the sinusoidal steady state. However, the algebra turns out to be unnecessarily complicated. We will thus follow instead a standard practice and consider a fictitious complex exponential excitation⁵⁷⁻⁵⁹ of the form

$$v_{gs}(t) = V_{gs}e^{j\omega t} \quad (9.4.49a)$$

$$v_{bs}(t) = V_{bs}e^{j\omega t} \quad (9.4.49b)$$

$$v_{ds}(t) = V_{ds}e^{j\omega t} \quad (9.4.49c)$$

where capital symbols with lowercase subscripts denote time-independent phasor quantities which can, in general, be complex, and ω is the angular frequency (in rad/s). Since the equations relating the various *small-signal* quantities (derived above) are linear, each small-signal quantity that results as an effect of the excitations in (9.4.49) in the steady state will also be equal to a complex quantity times $e^{j\omega t}$. In particular, we can write

$$u_i(x, t) = U_i(x, \omega)e^{j\omega t} \quad (9.4.50a)$$

$$i_i(x, t) = I_i(x, \omega)e^{j\omega t} \quad (9.4.50b)$$

$$i_d(t) = I_d(\omega)e^{j\omega t} \quad (9.4.50c)$$

$$i_g(t) = I_g(\omega)e^{j\omega t} \quad (9.4.50d)$$

$$i_b(t) = I_b(\omega)e^{j\omega t} \quad (9.4.50e)$$

Equations (9.4.50) are the various “responses” to the excitations (9.4.49). All these *complex* excitations and responses are fictitious, but they are useful for the following reason.⁵⁷⁻⁵⁹ If the *real* part of the above excitations is used instead to drive the device, all responses in the steady state will be given by the *real* part of the fictitious responses in (9.4.50). Now, the real part of any of the above excitations is a sinusoid. [For example, if M and ϕ are the magnitude and phase of V_{gs} , respectively, then the real part of $v_{gs}(t)$ in (9.4.49a) is simply $M \cos(\omega t + \phi)$.] Thus, working with the

above fictitious exponential functions provides all useful information about the actual sinusoidal steady state, with real excitations and real responses, only with greater mathematical ease. In addition, if the response to complex exponentials is known, the response to other types of waveforms can be determined by using transform techniques.⁵⁷⁻⁵⁹

The quantities in (9.4.49) and (9.4.50) can now be substituted into (9.4.42), (9.4.43), (9.4.40), (9.4.41), (9.4.44), (9.4.46), and (9.4.48). In all cases, $e^{j\omega t}$ appears as a common factor on both sides. Thus, we easily obtain (Prob. 9.14)

$$I_i(x, \omega) = - \frac{\mu W C'_{ox}}{\alpha_1} \frac{\partial}{\partial x} [U_i(x) U_i(x, \omega)] \quad (9.4.51a)$$

$$\frac{\partial I_i(x, \omega)}{\partial x} = - j\omega C'_{ox} W U_i(x, \omega) \quad (9.4.51b)$$

$$U_i(0, \omega) = V_{gs} + (\alpha_1 - 1)V_{bs} \quad (9.4.52a)$$

$$U_i(L, \omega) = (V_{gs} - V_{ds}) + (\alpha_1 - 1)(V_{bs} - V_{ds}) \quad (9.4.52b)$$

$$I_d(\omega) = I_i(L, \omega) \quad (9.4.53a)$$

$$I_g(\omega) = j\omega C'_{ox} W \left[L \frac{\alpha_1 - 1}{\alpha_1} (V_{gs} - V_{bs}) + \frac{1}{\alpha_1} \int_0^L U_i(x, \omega) dx \right] \quad (9.4.53b)$$

$$I_b(\omega) = j\omega(\alpha_1 - 1)C'_{ox} W \left[L \frac{1}{\alpha_1} (V_{bs} - V_{gs}) + \frac{1}{\alpha_1} \int_0^L U_i(x, \omega) dx \right] \quad (9.4.53c)$$

In the above equations, note the following: μ , W , L , and C'_{ox} are known device parameters, α_1 is known for a given bias V_{SB} from (9.4.1), and $U_i(x)$ is a known function of x , from (9.4.13). V_{gs} , V_{bs} , and V_{ds} are known phasors representing the excitation. Thus, for a given ω , (9.4.51) is a system of two differential equations in two unknown functions, $I_i(x, \omega)$ and $U_i(x, \omega)$. This system can be solved by using Bessel or Kelvin functions, with the boundary conditions given in (9.4.52);^{1,7,17,21,35,39,45,52,55} an alternate solution uses symbolic iterative techniques.²⁰ Once the functions $U_i(x, \omega)$ and $I_i(x, \omega)$ have been determined, they can be substituted in (9.4.53) to give $I_d(\omega)$, $I_g(\omega)$, and $I_b(\omega)$. The mathematical details are long (the procedure is outlined in the statement of Prob. 9.15) and will not be presented here. Below we summarize the form of the results. The final expressions are in the form

$$I_d(\omega) = \frac{N_{dd}(\omega)V_{ds} + N_{dg}(\omega)V_{gs} + N_{db}(\omega)V_{bs}}{D(\omega)} \quad (9.4.54)$$

$$I_g(\omega) = \frac{N_{gd}(\omega)V_{ds} + N_{gg}(\omega)V_{gs} + N_{gb}(\omega)V_{bs}}{D(\omega)} \quad (9.4.55)$$

$$I_b(\omega) = \frac{N_{bd}(\omega)V_{ds} + N_{bg}(\omega)V_{gs} + N_{bb}(\omega)V_{bs}}{D(\omega)} \quad (9.4.56)$$

where the quantities $N_{kl}(\omega)$ ($k, l = d, g, b$) and $D(\omega)$ are infinite series in $j\omega$:

$$N_{kl}(\omega) = n_{kl0} + (j\omega)n_{kl1} + (j\omega)^2n_{kl2} + \dots \quad (9.4.57a)$$

$$D(\omega) = d_0 + (j\omega)d_1 + (j\omega)^2d_2 + \dots \quad (9.4.57b)$$

The coefficients in these series up to second order as well as for all $N_{kl}(\omega)$ in (9.4.54) to (9.4.56), are given in Appendix N. One is finally able to find the y parameters, by comparing (9.4.54) to (9.4.56) with (9.3.8):

$$y_{dd} = \frac{N_{dd}(\omega)}{D(\omega)}, \quad y_{dg} = \frac{N_{dg}(\omega)}{D(\omega)}, \quad y_{db} = \frac{N_{db}(\omega)}{D(\omega)} \quad (9.4.58)$$

$$y_{gd} = \frac{N_{gd}(\omega)}{D(\omega)}, \quad y_{gg} = \frac{N_{gg}(\omega)}{D(\omega)}, \quad y_{gb} = \frac{N_{gb}(\omega)}{D(\omega)} \quad (9.4.59)$$

$$y_{bd} = \frac{N_{bd}(\omega)}{D(\omega)}, \quad y_{bg} = \frac{N_{bg}(\omega)}{D(\omega)}, \quad y_{bb} = \frac{N_{bb}(\omega)}{D(\omega)} \quad (9.4.60)$$

For example, using (9.4.57) in (9.4.59), we have

$$y_{gd} = \frac{n_{gd0} + (j\omega)n_{gd1} + (j\omega)^2n_{gd2} + \dots}{d_0 + (j\omega)d_1 + (j\omega)^2d_2 + \dots} \quad (9.4.61)$$

The y parameters can be computed, for a given frequency, to any desired accuracy by keeping an appropriate number of terms in the numerator and the denominator. The values thus obtained can be substituted in the small-signal equivalent circuit of Fig. 9.15.

Consider now the equivalent circuit in Fig. 9.17. In this circuit, only three of the parameters determined above appear directly: y_{gd} , y_{gb} , and y_{bd} . The rest of the parameters can be trivially found from (9.3.7) and (9.3.10):

$$y_{gs} = -y_{gg} - y_{gd} - y_{gb} \quad (9.4.62a)$$

$$y_{bs} = -y_{bb} - y_{bd} - y_{bg} \quad (9.4.62b)$$

$$y_{sd} = -y_{dd} - y_{gd} - y_{bd} \quad (9.4.62c)$$

$$y_m = y_{dg} - y_{gd} \quad (9.4.62d)$$

$$y_{mb} = y_{db} - y_{bd} \quad (9.4.62e)$$

$$y_{mx} = y_{bg} - y_{gb} \quad (9.4.62f)$$

where the quantities in the right-hand side are given by (9.4.58) to (9.4.60).

We will write the expressions for the parameters of the model in Fig. 9.17 in a way that will help relate the model to the one developed in Sec. 8.3. We start from (9.4.58) to (9.4.60). In each y -parameter expression, we factor out the first nonzero term of the numerator. For example, consider y_{gd} in (9.4.61). From Appendix N we have $n_{gd0} = 0$ and $d_0 = 1$. Thus, we can write:

$$y_{gd} = j\omega n_{gd1} \frac{1 + j\omega(n_{gd2}/n_{gd1}) + \dots}{1 + j\omega d_1 + \dots} \quad (9.4.63)$$

A look at Appendix N reveals that $-n_{gd1}$ has exactly the same expression as C_{gd} in (8.3.10). Thus we can write

$$y_{gd} = -j\omega C_{gd} \frac{1 + j\omega(n_{gd2}/n_{gd1}) + \dots}{1 + j\omega d_1 + \dots} \quad (9.4.64)$$

Proceeding in a similar manner, we can find expressions for all parameters in Fig. 9.17. As each is being developed, part of the expression can be recognized as a familiar small-signal quantity discussed in Chap. 8.† The results are summarized below (minus signs are used in front of the y parameters, corresponding to Fig. 9.17):

$$-y_{gs} = j\omega C_{gs} \frac{1 + j\omega\tau_2 + \dots}{1 + j\omega\tau_1 + \dots} \quad (9.4.65a)$$

$$-y_{bs} = j\omega C_{bs} \frac{1 + j\omega\tau_2 + \dots}{1 + j\omega\tau_1 + \dots} \quad (9.4.65b)$$

$$-y_{gd} = j\omega C_{gd} \frac{1 + j\omega\tau_3 + \dots}{1 + j\omega\tau_1 + \dots} \quad (9.4.65c)$$

$$-y_{bd} = j\omega C_{bd} \frac{1 + j\omega\tau_3 + \dots}{1 + j\omega\tau_1 + \dots} \quad (9.4.65d)$$

$$-y_{gb} = j\omega C_{gb} + \frac{(j\omega)^2 C_{gb, \text{sat}} \tau_4 + \dots}{1 + j\omega\tau_1 + \dots} \quad (9.4.65e)$$

$$-y_{sd} = \frac{g_{sd}}{1 + j\omega\tau_1 + \dots} \quad (9.4.65f)$$

$$y_m = \frac{g_m}{1 + j\omega\tau_1 + \dots} \quad (9.4.65g)$$

†Some of the expressions obtained will be valid with best accuracy at low V_{DS} values because they include α_1 , which can be traced to our initial, simplifying assumption in (9.4.1). This same restriction ($\alpha = \alpha_1$) was encountered at various points in Sec. 8.3, and was subsequently removed. This restriction will soon be removed from the present model also.

$$y_{mb} = \frac{g_{mb}}{1 + j\omega\tau_1 + \dots} \quad (9.4.65h)$$

$$y_{mx} = 0 \quad (9.4.65i)$$

where (after lengthy algebra)

$$\tau_1 = \frac{4}{15} \frac{1}{\omega_o} \frac{1 + 3\eta + \eta^2}{(1 + \eta)^3} \quad (9.4.66a)$$

$$\tau_2 = \frac{1}{15} \frac{1}{\omega_o} \frac{2 + 8\eta + 5\eta^2}{(1 + \eta)^2(1 + 2\eta)} \quad (9.4.66b)$$

$$\tau_3 = \frac{1}{15} \frac{1}{\omega_o} \frac{5 + 8\eta + 2\eta^2}{(1 + \eta)^2(2 + \eta)} \quad (9.4.66c)$$

$$\tau_4 = \frac{2}{15} \frac{1}{\omega_o} \frac{2 + 13\eta + 30\eta^2 + 13\eta^3 + 2\eta^4}{(1 + \eta)^5} \quad (9.4.66d)$$

with η as given by (4.5.38) and plotted in Fig. 4.20, and

$$\omega_o = \frac{\mu(V_{GS} - V_T)}{\alpha L^2} \quad (9.4.67)$$

We note that the numerators in (9.4.65f) to (9.4.65h) do not contain frequency-dependent terms. This is because of cancellations which occur when (9.4.62c) to (9.4.62e) are used. For $-y_{gb}$ in (9.4.65e), we did not use the form of the relations immediately above it because, for this parameter, such a form causes numerical difficulties at $\eta = 1$. In the formula given, $C_{gb,sat}$ is the value of C_{gb} in saturation (Sec. 8.3).

Consider now *low* frequencies, so that $\omega \ll \omega_o$. Then (9.4.65) gives $-y_{gs} \approx j\omega C_{gs}$, $-y_{bs} \approx j\omega C_{bs}$, $-y_{gd} \approx j\omega C_{gd}$, $-y_{bd} \approx j\omega C_{bd}$, $-y_{sd} \approx g_{sd}$, $y_m \approx g_m$, and $y_{mb} \approx g_{mb}$. If, in addition, we use† $-y_{gb} \approx j\omega C_{gb}$, and since $y_{mx} = 0$, the model in Fig. 9.17 reduces to the model in Fig. 8.17. Thus the same model results, starting from different premises.

The quantity η as it results in the above development is given by (4.5.38), where $V'_{DS} = (V_{GS} - V_T)/\alpha$ with $\alpha = \alpha_1$. This value for α is not the best one to use, as explained in Sec. 4.5.3 (unless V_{DS} is small). It simply appears here because of our initial simplifying assumption in (9.4.1), which was necessary in order to obtain manageable results.‡ It is desirable, however, that our present model be consistent

†This approximation will be discussed shortly.

‡The reader may want to try developing the model by starting from a different value of α in order to appreciate the ensuing problems.

with the corresponding dc model in (4.5.37), where the restriction that $\alpha = \alpha_1$ is not used. Thus we will remove this restriction from the present model, and we will allow V'_{DS} to have the same appropriate value as in the dc equations (see the discussion in Sec. 4.5.3).

Similar comments apply to other quantities in (9.4.65) to (9.4.67) (C_{bs} , g_{sd} , etc.). Thus, the approach we have followed leads to expressions for these quantities which are familiar from Chap. 8, only with $\alpha_1 - 1$ in place of a somewhat different quantity. [For example, we find $g_{mb} = (\alpha_1 - 1)g_m$, which agrees with (8.2.20a), but which is not accurate for large V_{DS} , as can be seen from the development of that equation.] Also, in the saturation region, the value of g_{sd} resulting from the above development is zero, since we have not taken channel length modulation into account. All such limitations can be removed by adopting the following approach: *Any quantity in (9.4.65) to (9.4.67), which has already been encountered in Chap. 8, will be assumed to have the value given in that chapter.* In this way, at low frequencies, our present model will reduce to the one in Sec. 8.3, not only in topology (Fig. 8.17) but also in the values for its elements. This is a very desirable property. It makes possible the incorporation of all the refinements known for low-frequency small-signal parameters into the present model. Thus, the model will be very well behaved at low frequencies (and will, of course, also provide useful results at frequencies where low-frequency models fail).

Using (8.3.15) and (9.4.65) we obtain

$$\frac{y_{bs}}{y_{gs}} \approx \frac{y_{bd}}{y_{gd}} \approx \frac{y_{mb}}{y_m} \approx \frac{dV_T}{dV_{SB}} = \alpha_1 - 1 \quad (9.4.68)$$

These relations hold with good accuracy if V_{DS} or V_{GS} is small and/or V_{SB} is large, in which case the depletion region is approximately uniform along the channel. In other cases, the relations hold with limited accuracy.

We now derive some very useful approximations for the relations in (9.4.65). Consider y_{gs} as an example. If the frequency of operation satisfies $\omega\tau_2 \ll 1$, we can write $1 + j\omega\tau_2 \approx 1/(1 - j\omega\tau_2)$. Using this in (9.4.65a) and neglecting high-order terms, we obtain $-y_{gs} \approx j\omega C_{gs}/[1 + j\omega(\tau_1 - \tau_2)]$; similarly for $-y_{bs}$, $-y_{gd}$, and $-y_{bd}$. The validity of these approximations will be considered shortly. Thus, we have

$$-y_{gs} \approx \frac{j\omega C_{gs}}{1 + j\omega(\tau_1 - \tau_2)}, \quad \omega\tau_2 \ll 1 \quad (9.4.69a)$$

$$-y_{bs} \approx \frac{j\omega C_{bs}}{1 + j\omega(\tau_1 - \tau_2)}, \quad \omega\tau_2 \ll 1 \quad (9.4.69b)$$

$$-y_{gd} \approx \frac{j\omega C_{gd}}{1 + j\omega(\tau_1 - \tau_3)}, \quad \omega\tau_3 \ll 1 \quad (9.4.69c)$$

$$-y_{bd} \approx \frac{j\omega C_{bd}}{1 + j\omega(\tau_1 - \tau_3)}, \quad \omega\tau_3 \ll 1 \quad (9.4.69d)$$

Consider now $-y_{gb}$ in (9.4.65e). We have

$$-y_{gb} = j\omega C_{gb} + y_a \quad (9.4.69e)$$

where, neglecting high-order terms, y_a is given by

$$y_a \approx (j\omega)^2 \frac{C_{gb,\text{sat}}\tau_4}{1 + j\omega\tau_1} \quad (9.4.69f)$$

In saturation, and at frequencies where the other approximations we have been making are accurate, y_a can be assumed zero with little penalty. In nonsaturation, and especially with very small V_{DS} , this term can be the dominant one in y_{gb} , but the magnitude of y_{gb} is then very small anyway. The small currents that can be contributed by it are almost invariably masked by other larger currents (e.g., those contributed by the extrinsic gate-substrate capacitance). Thus, y_a can be omitted for many applications (Prob. 9.17).

For the rest of the parameters in (9.4.65) we simply drop the high-order terms in the denominators:

$$-y_{sd} \approx \frac{g_{sd}}{1 + j\omega\tau_1}, \quad \omega\tau_1 \ll 1 \quad (9.4.69g)$$

$$y_m \approx \frac{g_m}{1 + j\omega\tau_1}, \quad \omega\tau_1 \ll 1 \quad (9.4.69h)$$

$$y_{mb} \approx \frac{g_{mb}}{1 + j\omega\tau_1}, \quad \omega\tau_1 \ll 1 \quad (9.4.69i)$$

$$y_{mx} = 0 \quad (9.4.69j)$$

The admittances in the right-hand side of (9.4.69a) to (9.4.69d) are of the general form $j\omega C/(1 + j\omega\tau)$. Figure 9.19a shows a simple circuit that realizes such an admittance. Figure 9.19b shows a circuit that realizes the admittance in the right-hand

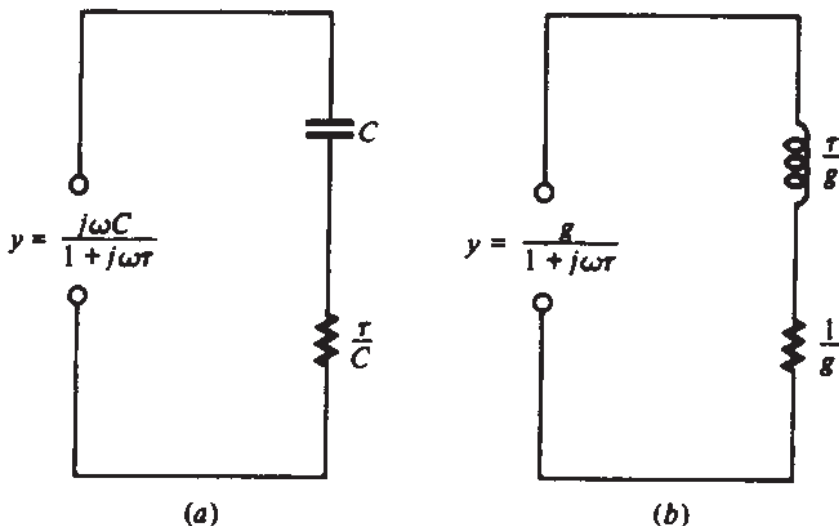
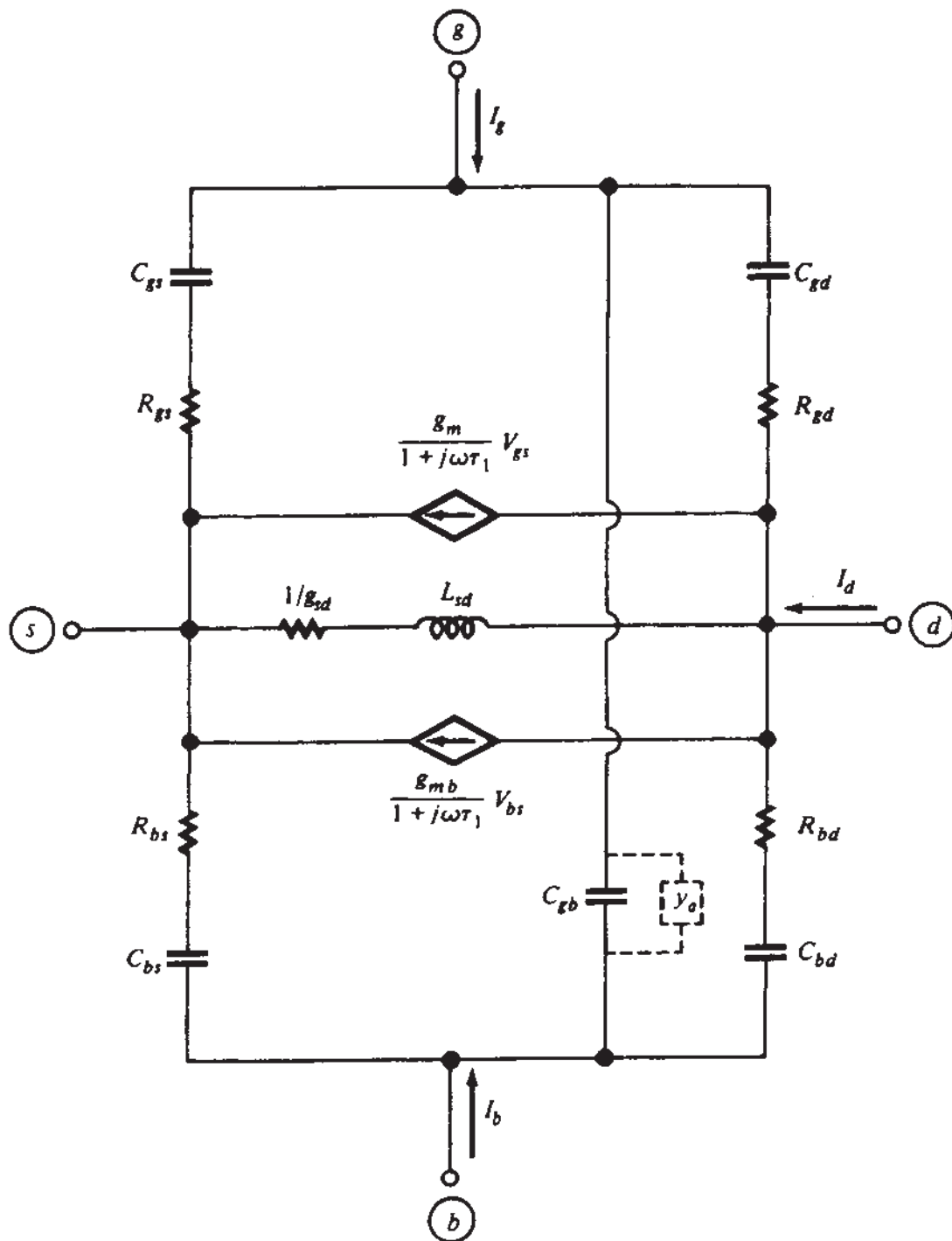


FIGURE 9.19

Simple circuits used to represent the admittances (a) in (9.4.69a) to (9.4.69d); (b) in (9.4.69g).

**FIGURE 9.20**

Small-signal non-quasi-static equivalent circuit model corresponding to (9.4.69).

side of (9.4.69g). These circuit representations in Fig. 9.19 can be verified through simple circuit analysis (Prob. 9.18).

With the help of Fig. 9.19, it is easy to see that using (9.4.69) the equivalent circuit of Fig. 9.17 takes the form shown in Fig. 9.20. The box shown in broken lines is y_a in (9.4.69f). In many applications this box can be omitted, for reasons already mentioned above. As follows from Fig. 9.19a and (9.4.69a) to (9.4.69d), we have

$$R_{gs}C_{gs} = R_{bs}C_{bs} = \tau_1 - \tau_2 \quad (9.4.70a)$$

$$R_{gd}C_{gd} = R_{bd}C_{bd} = \tau_1 - \tau_3 \quad (9.4.70b)$$

from which the resistance values can be calculated.† For the inductor we have, as follows from Fig. 9.19b and (9.4.69g),

$$L_{sd}g_{sd} = \tau_1 \quad (9.4.70c)$$

Plots for the resistances and the inductance are shown in Fig. 9.21.‡

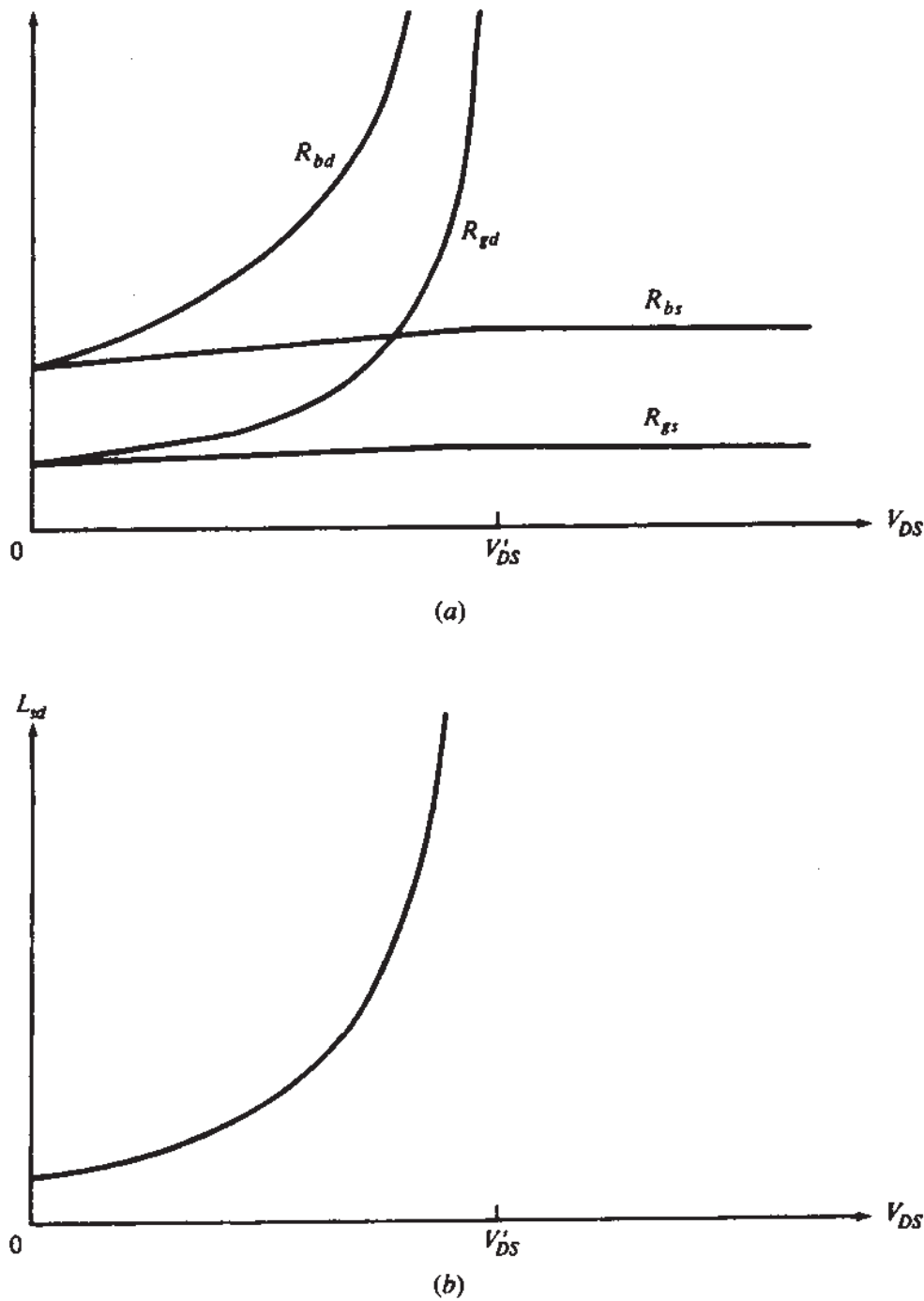
Despite the conditions of the form $\omega\tau_i \ll 1$ in (9.4.69), the model just presented is found satisfactory to about $\omega = \omega_o$. A comparison of the various models will be made in Sec. 9.4.3.

The various resistors and the inductor in Fig. 9.20 (in cooperation with the elements in series with them) can be viewed as representing some of the effects of the inversion layer's inertia when responding to rapid changes. In the following discussion, we assume that, in Fig. 9.18, only one small-signal voltage is nonzero at a time. If the source voltage is changing fast, the inversion layer will "hesitate" to respond, and the corresponding changes in the gate and substrate currents will lag behind the source voltage changes; this is modeled by R_{gs} , C_{gs} and R_{bs} , C_{bs} , respectively. The combinations R_{gd} , C_{gd} and R_{bd} , C_{bd} model the corresponding effects if the drain voltage is changing fast instead (in the nonsaturation region). The combination of L_{sd} and g_{sd} can be viewed as representing the inertia of the inversion layer in changing the source current when a fast-varying drain voltage demands such a change (in the nonsaturation region); this will be discussed shortly. Finally, the denominators in the values of the two current sources model the inertia of the inversion layer in changing the drain current when the gate or the substrate voltages are varying.

The appearance of an inductor in Fig. 9.20 may appear strange. To remove the "mystery" surrounding this element, consider the experiment shown in Fig. 9.22a. A transistor is biased in strong inversion with $V_{DS} = 0$, and a small signal is applied at the drain end of the channel. We chose $V_{DS} = 0$ to simplify things, since with such bias the channel can be thought of as a uniform resistor. The channel has a parasitic capacitance to the gate and the substrate. Although this capacitance, along with the channel resistance, forms a distributed RC element, to first order we can represent them with the lumped approximation shown in Fig. 9.22b. The capacitance C is shown grounded,

†Note that since the capacitances are proportional to $C_{ox} = C'_{ox}WL$ (Sec. 8.3), whereas τ_1 , τ_2 , and τ_3 are inversely proportional to ω_o [see (9.4.66) and (9.4.67)], the resistances will be proportional to $(\omega_o C_{ox})^{-1} = (L/W) \propto [\mu C'_{ox}(V_{GS} - V_T)]^{-1}$.

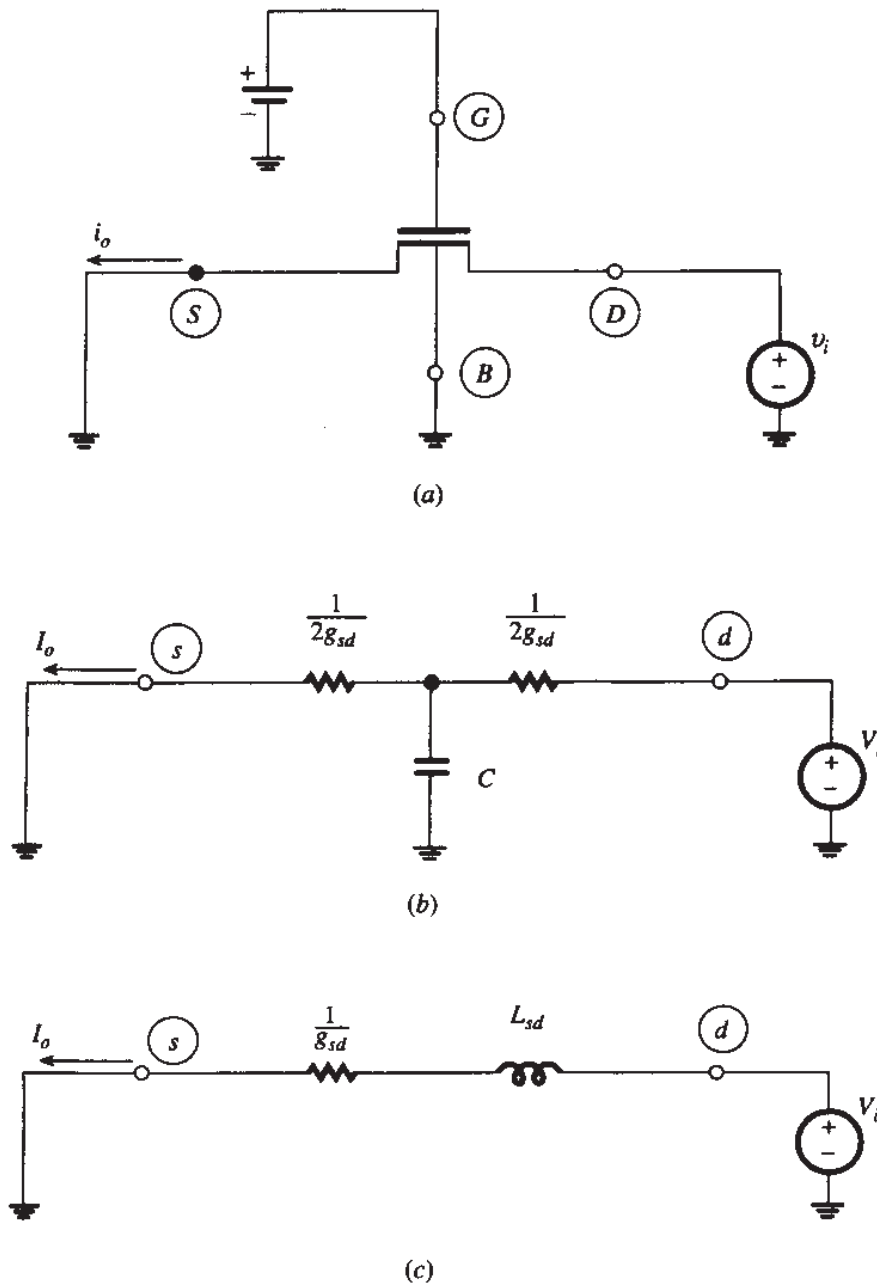
‡It is seen that R_{gd} , R_{bd} , and L_{sd} become infinite in saturation, just as do the impedances of the elements in series with them (assuming no channel length modulation). When the model is implemented as part of a CAD program, one should exercise care in order to avoid numerical difficulties.

**FIGURE 9.21**

Typical behavior of (a) the resistances R_{gs} , R_{gd} , R_{bs} , and R_{bd} ; (b) the inductance L_{sd} .

since both the gate and the substrate capacitance in (a) are connected to fixed voltages (i.e., “small-signal ground”). The proper value to use for C in this model is less than the total capacitance, since the capacitance close to the ends of the channel is practically shunted out (Prob. 9.26). Let us consider the small-signal current exiting from the source, shown in Fig. 9.22a. This current, and the driving voltage, are represented by complex phasors^{57–59} in Fig. 9.22b. Analysis of this circuit gives

$$I_o = \frac{g_{sd}}{1 + j\omega\tau} V_i \quad (9.4.71)$$

**FIGURE 9.22**

(a) A transistor with zero drain-source bias, and with a small voltage signal at the drain terminal. (b) A lumped model for modeling the current exiting from the source terminal. (c) Part of the model in Fig. 9.20, relevant to modeling the same current.

where $\tau = C/(4g_{sd})$. Let us now, instead of using the circuit in Fig. 9.22b, apply the small-signal equivalent circuit of Fig. 9.20 to the case of Fig. 9.22a. The result is shown in Fig. 9.22c, where branches shunted by the voltage sources or by short circuits are omitted, since they do not affect the outcome. The controlled sources are not shown, since at the bias point $V_{DS} = 0$ we have $g_m = g_{mb} = 0$. It is easily seen that an analysis of this circuit gives exactly (9.4.71) again, with $\tau = L_{sd}g_{sd}$ (see Prob. 9.26 for a quantitative comparison). Thus, the presence of the inductor in the circuit in Fig. 9.22c is no mystery: this circuit is just an equivalent circuit that produces the same

results for I_o as the circuit in Fig. 9.22b. Notice that the series RC circuits in Fig. 9.20 cannot model the effect we are discussing in relation to Fig. 9.22a, since those circuits are shunted by voltage sources and short circuits, and do not affect the value of I_o ; it is the presence of the inductor that provides the proper modeling in this case.† Notice from (9.4.71) that the higher the frequency, the smaller the magnitude of I_o , and the larger its phase lag with respect to V_i . This is what one would expect by considering the “inertia” of the channel to respond to the variations of $v_i(t)$. In fact, *small-signal non-quasi-static effects and distributed effects are two ways of looking at the same thing.*

It is clear that, as frequency is increased further and further, the representation of distributed effects by the lumped approximation in Fig. 9.22b will eventually fail; thus, the model in Fig. 9.22c will fail, too, and so will the model of Fig. 9.20 in general. One can extend the frequency limit of validity with higher-order (but more complicated) models,^{39,44,45,48–50,52,55} or by using several sections as in Fig. 7.12.

If the frequency of the small-signal voltages applied to the terminals becomes low enough, the inertia mentioned becomes negligible. Indeed, with decreasing frequency the magnitude of the impedances of C_{gs} , C_{bs} , C_{gd} , and C_{bd} increases. Relative to them, the series resistances in Fig. 9.20 become unimportant and can be omitted. Also, with decreasing frequency the impedance of the inductor decreases, and this element can be omitted in comparison to the resistance in series with it. Finally, the denominators in the current source values become approximately 1 at low frequencies. In this case, the model in Fig. 9.20 is thus seen to reduce to the model of Fig. 8.17.

The model can also be related^{35,41} to the complete quasi-static model of Sec. 9.2. At low frequencies, the series RC combinations in Fig. 9.20 reduce to the corresponding capacitances in Fig. 9.5, for the reasons discussed above. In addition, assuming $\omega\tau_1 \ll 1$, we can use the approximation $1/(1 + j\omega\tau_1) \approx 1 - j\omega\tau_1$, and we can write (9.4.69g) to (9.4.69i) as follows:

$$-y_{sd} \approx g_{sd} - j\omega\tau_1 g_{sd}, \quad \omega\tau_1 \ll 1 \quad (9.4.72a)$$

$$y_m \approx g_m - j\omega\tau_1 g_m, \quad \omega\tau_1 \ll 1 \quad (9.4.72b)$$

$$y_{mb} \approx g_{mb} - j\omega\tau_1 g_{mb}, \quad \omega\tau_1 \ll 1 \quad (9.4.72c)$$

A comparison with $-y_{sd}$, y_m , and y_{mb} for the complete quasi-static model [(9.3.11f) to (9.3.11h)] shows that the form is the same (recall that C_{sd} was a *nega-*

†The presence of the inductor in the model, in association with the capacitances in it, might suggest the possibility of “overvoltages,” associated with series RLC circuits.^{57,58} This can, in fact, be observed to some extent. Consider a transistor with the drain not connected to anything (and thus with $I_{DS} = 0$ and $V_{DS} = 0$), and driven by a bias and small-signal voltage between gate and source. In this case the channel is uniform, and for small signals it behaves as a passive resistance, with associated distributed capacitance to the gate. Such a device can exhibit, under certain conditions, an ac drain voltage somewhat larger than the input signal for a range of frequencies. This effect can be predicted by the model of Fig. 9.20 (Prob. 9.27) and can be verified by measurements. The effect can even be predicted by using a simple, passive lumped RC model of the distributed RC behavior of the channel, using a few resistors and capacitors (certain passive RC circuits can exhibit a voltage gain somewhat larger than unity, although they cannot exhibit power gain). The interested reader is referred to Prob. 9.27.

tive quantity). Furthermore an examination of the expressions for the various parameters in these equations reveals that $\tau_1 g_{sd} = -C_{sd}$, $\tau_1 g_m = C_m$, and $\tau_1 g_{mb} = C_{mb}$; therefore, (9.4.72a) to (9.4.72c) are in fact *identical* to (9.3.11f) to (9.3.11h), not only in form but even in *value* (Prob. 9.19). Thus, the model of Fig. 9.20 reduces to the complete quasi-static model of Fig. 9.5, assuming C_{mx} is negligible. Note in particular that the series resistance-inductance combination of Fig. 9.20 reduces to a parallel combination of a resistance and a *negative* capacitance. With further reduction in frequency, the terms containing ω in $-y_{sd}$, y_m , and y_{mb} become negligible, and the model reduces to the simple model of Fig. 8.17.

The coefficients of the controlled sources in Fig. 9.20 are complex. This might make it impossible to use this model *directly* in some computer analysis programs. The problem is circumvented by noting that we can write

$$\frac{g_m}{1 + j\omega\tau_1} V_{gs} = g_m V_1 \quad (9.4.73a)$$

$$\frac{g_{mb}}{1 + j\omega\tau_1} V_{bs} = g_{mb} V_2 \quad (9.4.73b)$$

where

$$V_1 = \frac{1}{1 + j\omega\tau_1} V_{gs} \quad (9.4.74a)$$

$$V_2 = \frac{1}{1 + j\omega\tau_1} V_{bs} \quad (9.4.74b)$$

and that it is very easy to develop V_1 from V_{gs} and V_2 from V_{bs} by using two simple circuits. This idea is exploited in Fig. 9.23, where it is easy to verify that (9.4.74) holds as long as $R_1 C_1 = \tau_1$ and $R_2 C_2 = \tau_1$.† However, in order not to upset the model, we have to make sure that the new elements added draw a negligible current (in comparison to the R_{gs} - C_{gs} and R_{bs} - C_{bs} combinations). This can be ensured, for example, by using

$$C_1 = 0.001 C_{gs} \quad (9.4.75a)$$

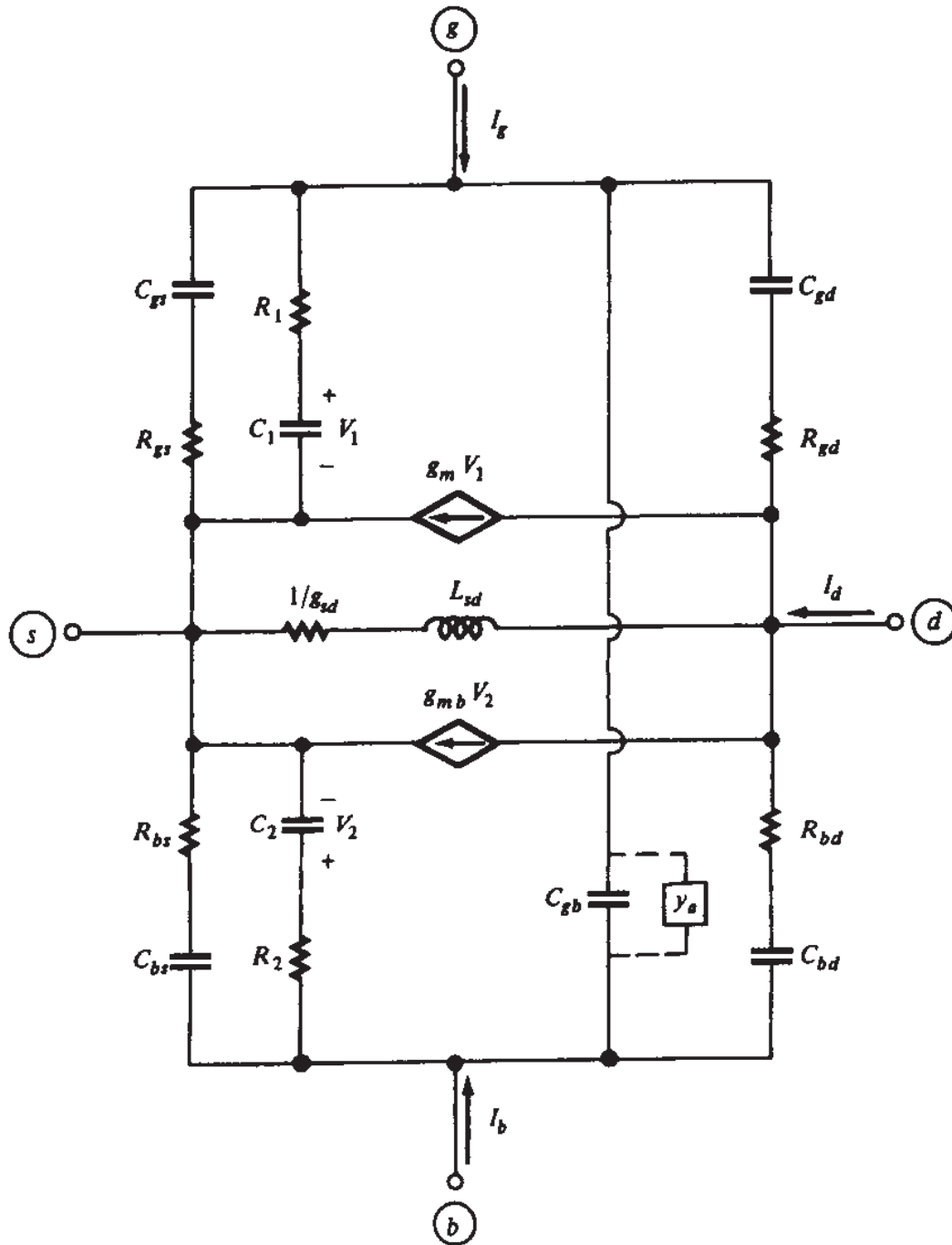
$$R_1 = \frac{\tau_1}{C_1} \quad (9.4.75b)$$

$$C_2 = 0.001 C_{bs} \quad (9.4.75c)$$

$$R_2 = \frac{\tau_1}{C_2} \quad (9.4.75d)$$

Similarly, it is possible to replace y_a (if it is decided to include this element in the model) by a combination of circuit elements with real values.

†Unfortunately, we could not have used the voltages across C_{gs} and C_{bs} themselves, since $R_{gs}C_{gs} = R_{bs}C_{bs} = \tau_1 - \tau_2$ instead of being equal to the desired τ_1 .

**FIGURE 9.23**

The model of Fig. 9.20 modified to avoid complex coefficients in the controlled-current sources.

In the saturation region, the model of Fig. 9.20 assumes a simple form, easily derived from the relations given above, and shown in Fig. 9.24. In this region g_{sd} models channel length modulation (Sec. 6.2). This effect has largely been studied at low frequencies, and was not included in the derivations of the present model. Thus, there is no reason to assume that L_{sd} , as results from (9.4.70c), will be the correct value to use in the saturation region. In fact, in saturation models, L_{sd} in Fig. 9.24 is often replaced by a short circuit. This is not likely to be a problem in practice, since

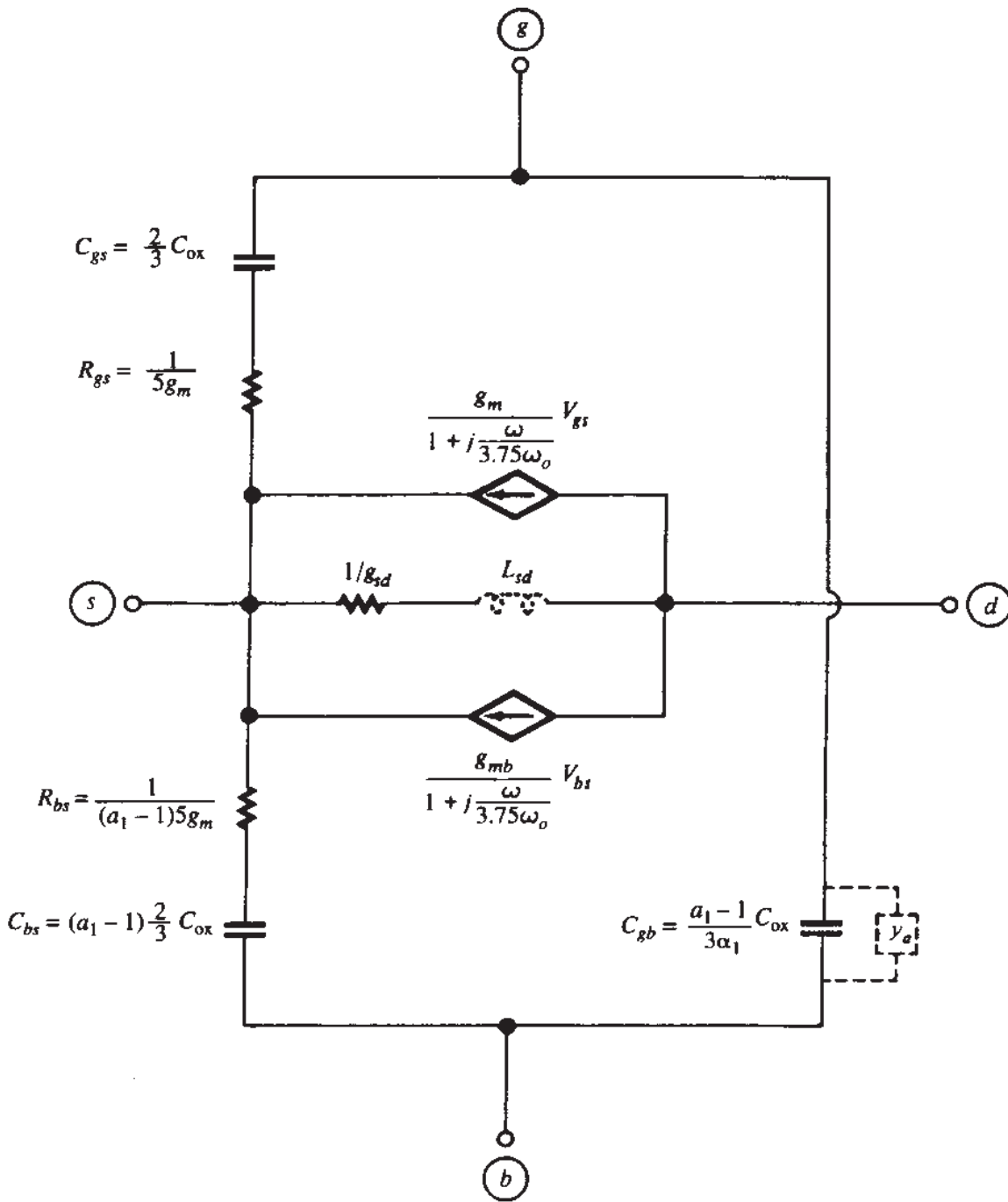


FIGURE 9.24

The model of Fig. 9.20 for the special case of operation in the saturation region.

in a circuit capacitances invariably exist between drain and source (e.g., because of another transistor connected to the one under consideration, or parasitics). At high frequencies, where the impedance of L_{sd} could become comparable to $1/g_{sd}$, the impedance of the above capacitances is low, and relatively large small-signal currents can flow in them. Compared to these, the minute small-signal current flowing through $1/g_{sd}$ and L_{sd} is likely to be negligible.

9.4.3 Other Approximations and Higher-Order Models

The model we have presented is valid up to about $\omega = \omega_o$, as already mentioned. In developing it, we used a certain type of approximation, e.g., (9.4.65a) was approximated by (9.4.69a). One may wonder why not simply drop the high-order terms in (9.4.65a), and use $-y_{gs} = j\omega C_{gs}(1 + j\omega\tau_2)/(1 + j\omega\tau_1)$ without any further manipulation. The answer is that such an approach would not be wise for two reasons: (1) The region of validity for a model based on this type of approximation turns out to be about the same as for the one we presented, whereas the equivalent circuit becomes more complicated; and (2) the degradation of such a model with frequency would not be “graceful.” This means that, as the frequency is increased above the limit of validity, the error in some parameters (especially in their phase) soon would become quite large. In contrast to this, the simple model we have presented degrades “gracefully,” with the errors increasing rather mildly as the frequency limit of validity is exceeded. The explanation for this somewhat unexpected behavior is as follows. Once the frequency is high enough for the missing high-order terms to become important, the new approximation we are considering is inadequate and large errors can result, since the numerator is of second order whereas the denominator is of first order. In (9.4.69a), the approximation made turns out to be in such a direction as to partially compensate for the effect of the missing terms; hence the more graceful degradation. Graceful degradation of models is important (although one is not supposed to use models outside the specified limits of validity). For example, a user might not be aware of such limits (which is usually the case in CAD) or might have to use a model outside the region of validity, just because there is no better model available. In such cases, it is desirable that the resulting error not be excessive. A good CAD program, though, should warn the user when limits of validity are exceeded.

Models valid up to frequencies higher than ω_o can be developed by keeping an appropriate number of high-order terms in the y -parameter expressions.^{39,40,43–45,48–50,52,55} This must be done with care, to ensure graceful degradation in the sense discussed above. One might wonder what need there could be for models valid at such high frequencies when the model of Fig. 9.20 is capable of predicting device performance all the way to ω_o , which is near the intrinsic transition frequency of the device (8.3.31). Consider, however, a circuit with devices operating well below *their* ω_o . In the same circuit there may be *other* devices with *longer* channels for which ω_o is much *smaller* [see (9.4.67)]. Such devices might be providing useful functions other than gain. Unless these devices are modeled correctly, the predicted performance for the whole circuit might be in error.

In such cases, an alternative to using very complicated high-order models may be considered for the devices with the longer channels. Let ω_{highest} be the highest operating frequency of interest for a given circuit. Evaluate ω_o for all relevant devices in the circuit. Those with $\omega_o > \omega_{\text{highest}}$ can be modeled as discussed in this section. The others can be divided into “subtransistors” (Fig. 7.12), such that ω_o of each subtransistor is larger than ω_{highest} . The subtransistors can then also be modeled as discussed in this section. Notice that, even though the internal subtransistors may have a small

channel length, these subdevices do not have real sources and drains; thus, the model used for them should be free of short-channel effects.

The model we have derived in this section is valid in strong inversion only. It is possible to derive a non-quasi-static model valid in all regions of operation, starting from the general charge sheet model concepts discussed in Sec. 4.3.³⁹ The topology of such a model is exactly the same as in Fig. 9.20, but, of course, the expressions for the model parameters are different. Unfortunately, these expressions are complicated.

9.4.4 Model Comparison

Figure 9.25 shows plots³⁹ of $|y_m|/g_m$ and of the phase of y_m versus frequency (log scale) in strong inversion, for $\eta = 0.5$ ($V_{DS} = V'_{DS}/2$). The frequency ω_o is given by (8.3.6), repeated here for convenience:

$$\omega_o = \frac{\mu(V_{GS} - V_T)}{\alpha L^2} \quad (9.4.76)$$

In each plot, curve *a* is for the simple model of Fig. 8.17, curve *b* is for the complete quasi-static model of Fig. 9.5, and curve *c* is for the model of Fig. 9.20. Finally, curve *d* is for a model resulting by keeping many terms in the numerators and denominators

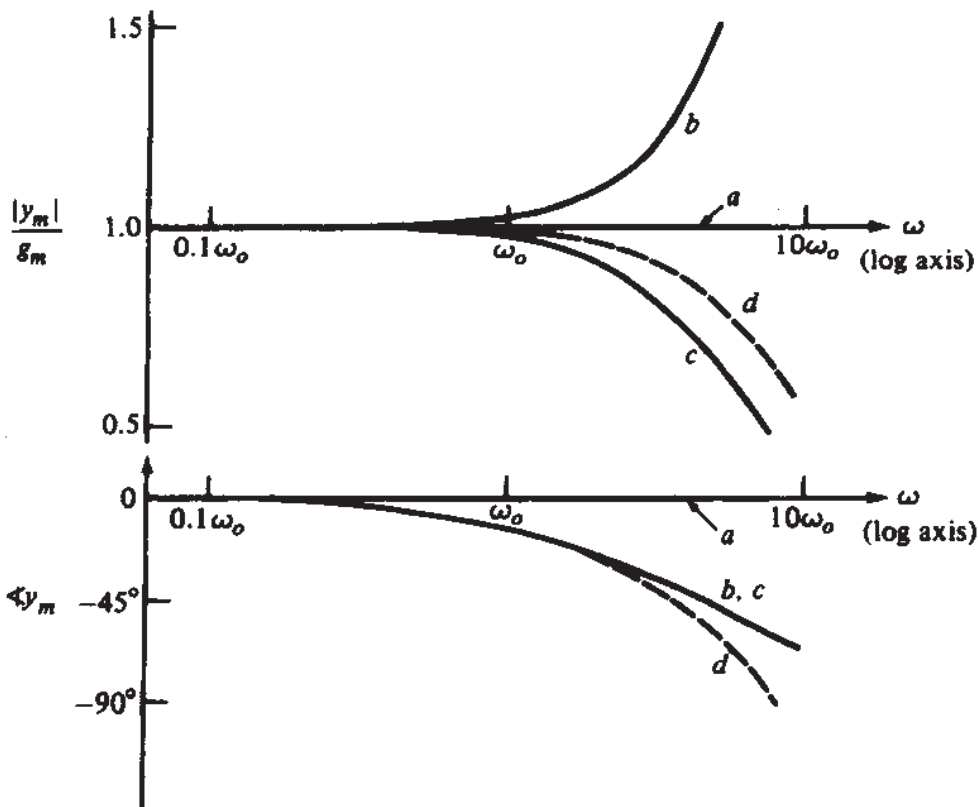


FIGURE 9.25

Normalized magnitude and phase of y_m versus frequency for a transistor operating in nonsaturation with $\eta = 0.5$.³⁹ (a) Simple model of Fig. 8.17; (b) complete quasi-static model of Fig. 9.5; (c) model of Fig. 9.20; (d) numerical result (© 1985 by IEEE).

of (9.4.65). This model is valid even beyond $10\omega_o$.³⁹ It is seen in Fig. 9.25b that going from a to c produces a drastic improvement in the region of validity. The region of validity for b is limited by the fact that, at high frequencies the error in the magnitude becomes severe. This is because y_m contains a right-half-plane zero for this model (9.3.11g) in contrast to the left-half-plane pole in y_m for c (9.4.69h). The upward-going magnitude for b at high frequencies is clearly unrealistic, since it suggests an enhancement in the forward gate-to-drain action, contrary to one's expectation that, at high frequencies, control of the gate on the drain current is gradually lost due to the inversion layer's inertia. In fact, as far as magnitude is concerned, a is better than b , although it corresponds to a simpler model! Going to c eliminates the problem of b , and now both magnitude and phase are predicted satisfactorily up to about ω_o .

Similar plots can be constructed for other operating points and for other parameters in an attempt to compare the various models and determine upper frequency limits of validity.³⁹ It is found that the upper frequency limit of validity for a parameter of a given model depends on which parameter it is, what is the operating point, what accuracy is desired, whether magnitude or phase is of most interest, etc. Furthermore, one can always construct pathological cases where a model will fail in some way (Prob. 9.17). Thus, the frequency limits of validity we have been providing should be considered in the above light. They are only rough indications of the regions within which a given model will perform satisfactorily in most cases. To summarize, the frequency limits of validity we have suggested for strong inversion models in this and the previous chapters are as follows:

- | | |
|--|-----------------------|
| 1. Quasi-static model without transcapacitors (Fig. 8.17): | $\frac{\omega_o}{10}$ |
| 2. Quasi-static model with transcapacitors (Fig. 9.5): | $\frac{\omega_o}{3}$ |
| 3. First-order non-quasi-static model (Fig. 9.20): | ω_o |

9.5 HIGH-FREQUENCY NOISE

Non-quasi-static effects influence the power spectral density of the drain current noise at very high frequencies.⁶⁰⁻⁶⁹ It turns out, though, that this happens mainly at frequencies higher than ω_o , which is the limit of validity of our first-order non-quasi-static model. Thus, the drain current power spectral density in the model of Fig. 8.39, can be calculated as in Sec. 8.5. There is, however, another effect that should be considered at high frequencies, as explained below.

Thermal noise in strong inversion was seen in Sec. 8.5 to be the result of random potential fluctuations in the channel. These fluctuations are coupled to the gate terminal through the oxide capacitance, and cause a gate noise current to flow even if all terminal voltages are fixed.^{60-69,45} This current is called *induced gate noise*. At high frequencies, the impedance of the gate capacitance becomes smaller, and this effect becomes more pronounced. Thus, non-quasi-static models, being appropriate for high-frequency work, should include a model for this noise.

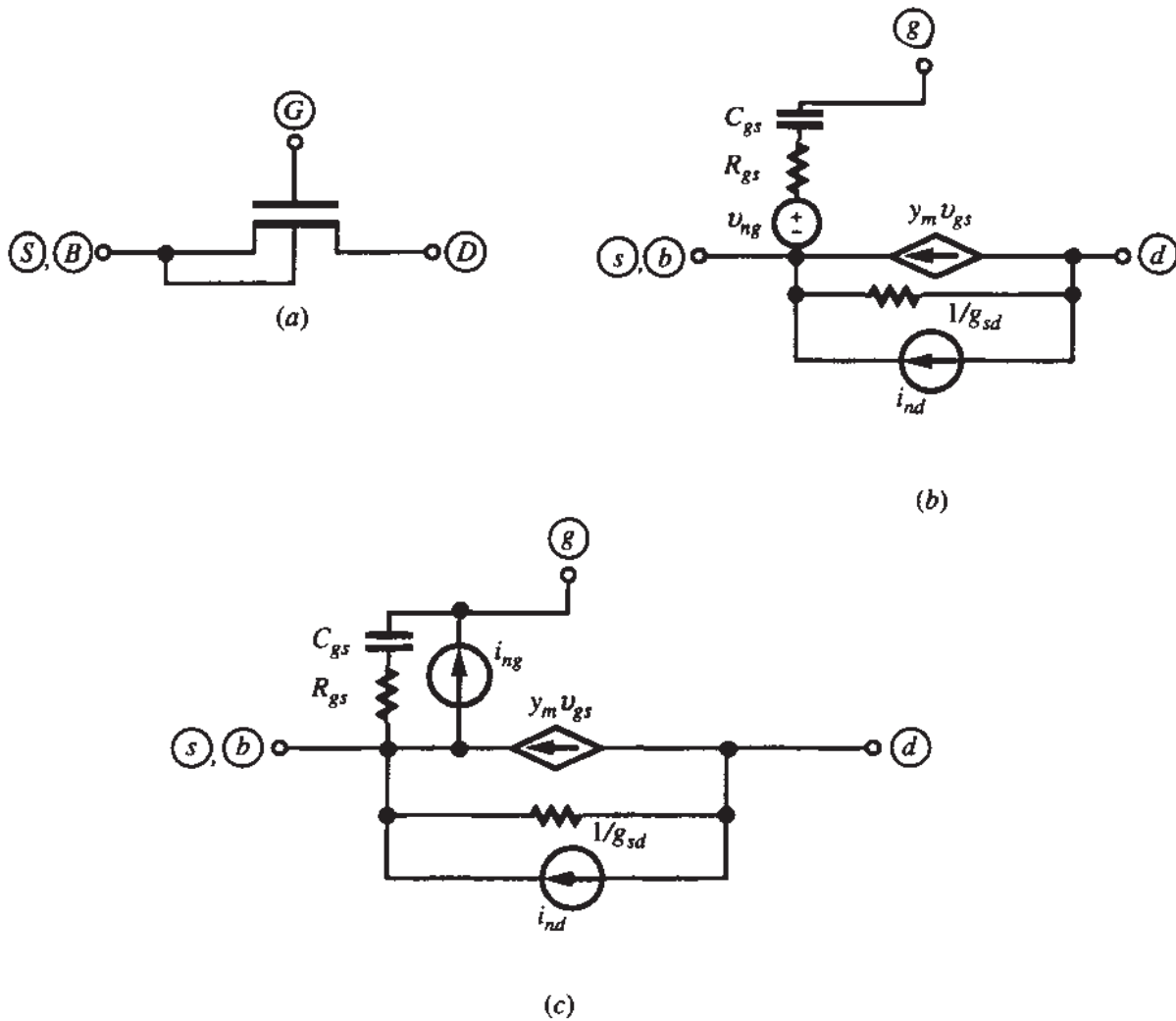


FIGURE 9.26

(a) A transistor with substrate shorted to source, assumed to operate in the saturation region. (b) Equivalent small-signal circuit for (a), including noise sources. (c) Alternative representation of the circuit in (b).

To obtain a feel for this effect, consider, as an example, a transistor in the strong-inversion saturation region, connected in a common-source configuration as shown in Fig. 9.26a, and ignore extrinsic elements. The corresponding intrinsic small-signal model can be taken from Fig. 9.24, and is given in Fig. 9.26b, with only the most important elements shown. To model the drain noise current, a noise current source i_{nd} in the drain port is used as discussed in Sec. 8.5. Consider now the resistance R_{gs} . The presence of this element in the model has been seen before to occur due to the resistance of the channel, which causes distributed RC effects in conjunction with the gate capacitance. If a complete non-quasi-static analysis is carried out for the influence of such effects on the induced gate noise,^{60,61,64–67} a surprisingly simple, and by no means obvious, result is found: it turns out that the induced gate current can be estimated by assuming that R_{gs} produces the same amount of thermal noise an actual resistor of value R_{gs} would produce.⁶³ This noise is modeled in Fig. 9.26b as a noise voltage source v_{ng} . According to (8.5.3), the power spectral density of this noise will be $S_{vng} = 4kTR_{gs}$. Various accurate (and very complicated) calculations^{60,61,64–67} give

practically the same result, within a factor close to unity. According to the results in Ref. 67, this factor is 4/3; thus we have

$$S_{vng} = 4kT \left(\frac{4}{3} R_{gs} \right), \quad \text{saturation} \quad (9.5.1)$$

Another common representation for the induced gate noise uses a "Norton equivalent" of the gate branch in Fig. 9.26b.⁶³ Such a circuit can be found by replacing the series voltage source by a parallel current source, as shown in Fig. 9.26c. From classical Thevenin-to-Norton conversion⁵⁷⁻⁵⁹ it can be concluded that the power spectral density S_{ing} of this source will be S_{vng} times the square of the magnitude of the admittance of the $C_{gs} - R_{gs}$ branch (the square comes from the fact that power spectral densities are associated with mean square values; see Sec. 8.5). From the value given in Fig. 9.24, it is easily seen that for frequencies below ω_o the admittance of this branch is dominated by C_{gs} , and its magnitude is thus approximately given by ωC_{gs} ; thus

$$S_{ing} = 4kT \left(\frac{4}{3} R_{gs} \right) \omega^2 C_{gs}^2, \quad \text{saturation,} \quad \omega < \omega_o \quad (9.5.2)$$

Using for R_{gs} and C_{gs} the saturation values given in Fig. 9.24, with $g_m = (W/L)\mu C'_{ox}(V_{GS} - V_T)$ (assuming $\alpha \approx 1$) the above equation becomes (Prob. 9.28)

$$S_{ing} = 4kT \frac{\omega^2 (C'_{ox} WL)^2}{\frac{W}{L} \mu C'_{ox} (V_{GS} - V_T)} \frac{16}{135}, \quad \text{saturation,} \quad \omega < \omega_o \quad (9.5.3)$$

This result agrees exactly with those obtained using very complicated calculations, in which infinitesimal elements of the channel are considered, their contribution to the current noise is found, and their combined effect is calculated.^{60-67,45} In fact, this development can be carried out directly in the nonsaturation region, giving an equation of the same form:^{61,66}

$$S_{ing} = 4kT \frac{\omega^2 (C'_{ox} WL)^2}{\frac{W}{L} \mu C'_{ox} (V_{GS} - V_T)} K_2(V_{DS}), \quad \omega < \omega_o \quad (9.5.4)$$

where the factor $K_2(V_{DS})$ is $1/12 \approx 0.083$ at $V_{DS} = 0$, and gradually rises to $16/135 \approx 0.12$ in saturation;⁶⁶ however, in practice $K_2(V_{DS})$ will rise to values *larger* than that (e.g., by a factor of 2 to 5), due to hot electron effects⁶⁹ (Sec. 8.5.4).

Similar calculations can be used to determine the degree of correlation between the gate and drain thermal-induced noise currents (a partial correlation is to be expected, since both currents have the same physical origin). In circuit calculations, caution is needed to take this correlation properly into account. One way to do this is to characterize the correlation by what is called the "cross spectral density" (a quantity analogous to the power spectral density, but defined for two noise currents^{66,67}).

For readers familiar with circuit noise calculations, we mention that the cross spectral density of the gate and drain thermal-induced currents in saturation is given by†

$$S_{ig,id} = 4kT \frac{1}{6} j\omega C_{gs}, \quad \omega < \omega_o \quad (9.5.5)$$

$S_{ig,id}$ is found in some treatments^{61,66} to decrease toward 0 as V_{DS} is reduced toward 0; in other treatments, the value of this quantity is taken to be roughly the same as in saturation.⁶⁸

9.6 CONSIDERATIONS IN MOSFET MODELING FOR RF APPLICATIONS

MODEL TOPOLOGIES. The frequencies used for radio communications cover a wide range. Usually, radio-frequency (RF) modeling of MOSFETs refers to modeling at frequencies up to the transition (cutoff) frequency of the devices. For such applications, small-signal non-quasi-static modeling becomes indispensable for the intrinsic part, and Secs. 9.3 and 9.4 are of direct relevance. However, the extrinsic part must be considered, too. Distributed effects in that part can be modeled using lumped approximations. An example is shown in Fig. 9.27a (noise sources are not shown for simplicity). A non-quasi-static model is used for the intrinsic part; this makes sense, since both long- and short-channel devices have been found to exhibit non-quasi-static effects.⁵¹ The resistances are those of the extrinsic gate, substrate, and source-drain regions. As seen, each is divided into subresistances connected to a common node, so that the appropriate capacitances can be connected to that node. For example, the elements R_{se1} , R_{se2} , and C_{bse} model, to first order, what is actually a distributed RC effect of the source resistance and its capacitance to the substrate. For noise modeling, each resistance can be modeled as in Fig. 8.32, and of course intrinsic noise sources for the intrinsic part should be used as discussed in the previous section.

Although the level of modeling exhibited by Fig. 9.27a is very desirable, in practice it is difficult to use because the individual values of the subresistances are difficult to determine. This is particularly true for the gate and substrate subresistances. In such cases, one may have to resort to the model shown in Fig. 9.27b. At times, models even simpler than that are used.⁷⁰⁻⁸⁰ An example is shown in Fig. 9.28. Figure 9.28a shows a transistor with the source and substrate shorted together (a very common connection). For operation in the saturation region, it is claimed in the literature that the model in Fig. 9.28b is valid.‡ The intrinsic part of the model is taken from Fig. 9.24. Note, however, that this model cannot be derived from that in Fig. 9.27 since, although the external source and substrate terminal are shorted together,

†Equivalently, one can give the (cross) correlation coefficient, defined^{66,67} as $\rho = S_{ig,id} / \sqrt{S_{ig} S_{id}}$. From the results given here, the value of this quantity in saturation can be found to be $0.395j$.

‡Sometimes inductances are added in series with the terminals to model the leads of discrete transistors, or even the connections to integrated transistors.

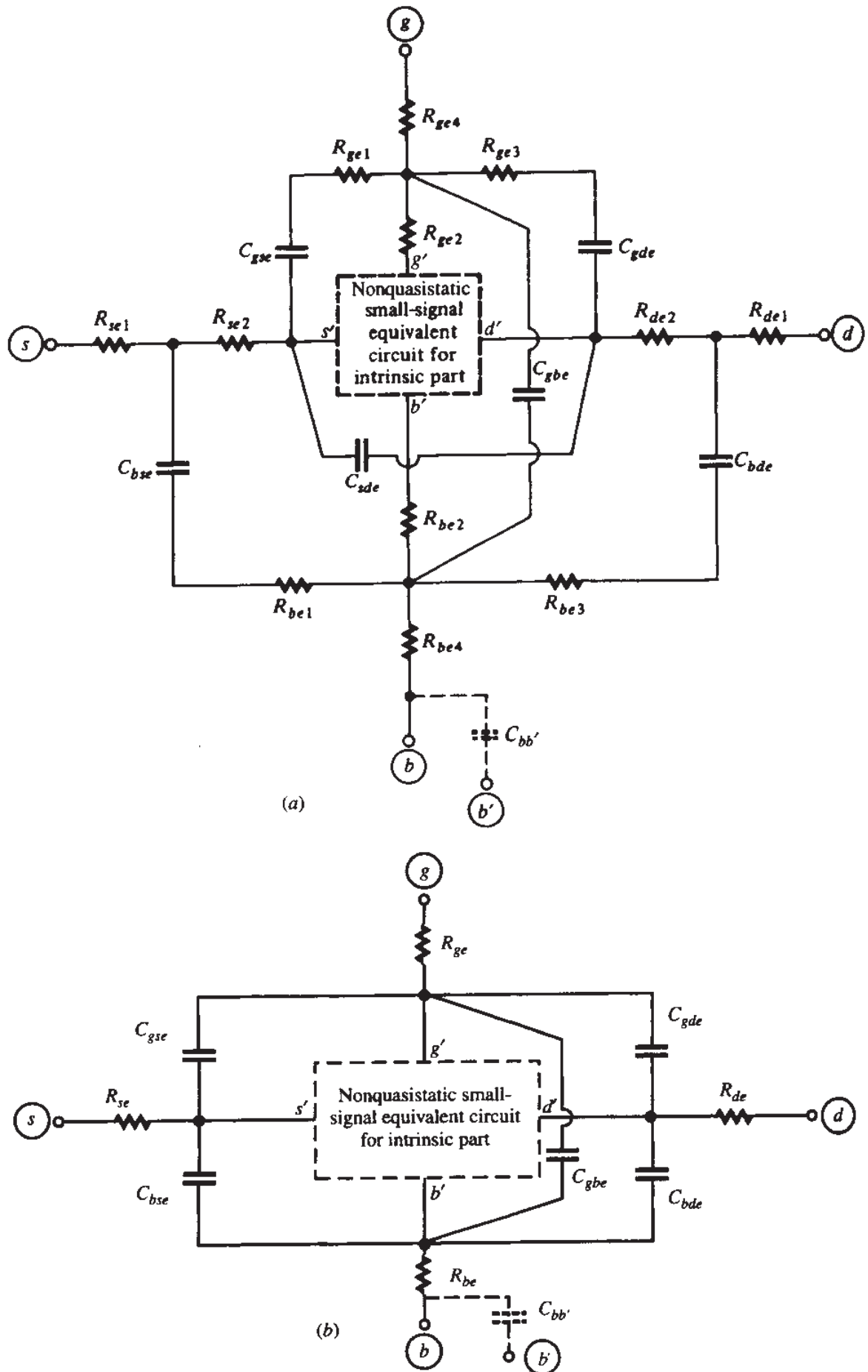
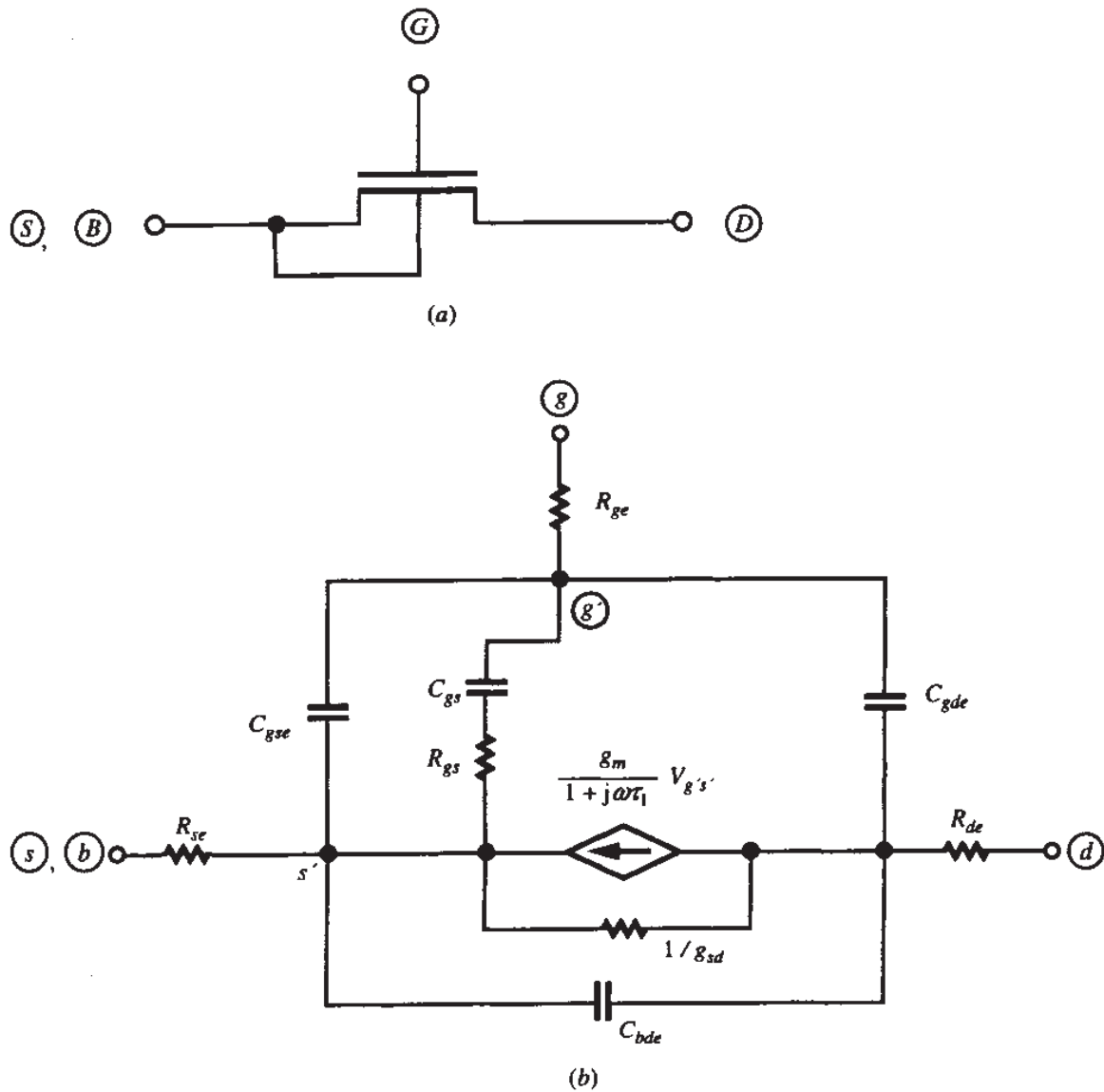


FIGURE 9.27
Small-signal models for the complete transistor: (a) is more accurate; (b) is more practical.

**FIGURE 9.28**

(a) A transistor with substrate shorted to source. (b) Small-signal model sometimes used for the connection in (a) in the saturation region.

R_{se} and R_{be} prevent the *intrinsic* source and substrate terminals (s' and b' in Fig. 9.27b) from being short-circuited. At times, an even simpler model is used; for example, R_{gs} and/or the frequency dependence of the gate transadmittance are omitted, or the intrinsic device is modeled using the complete quasi-static model.

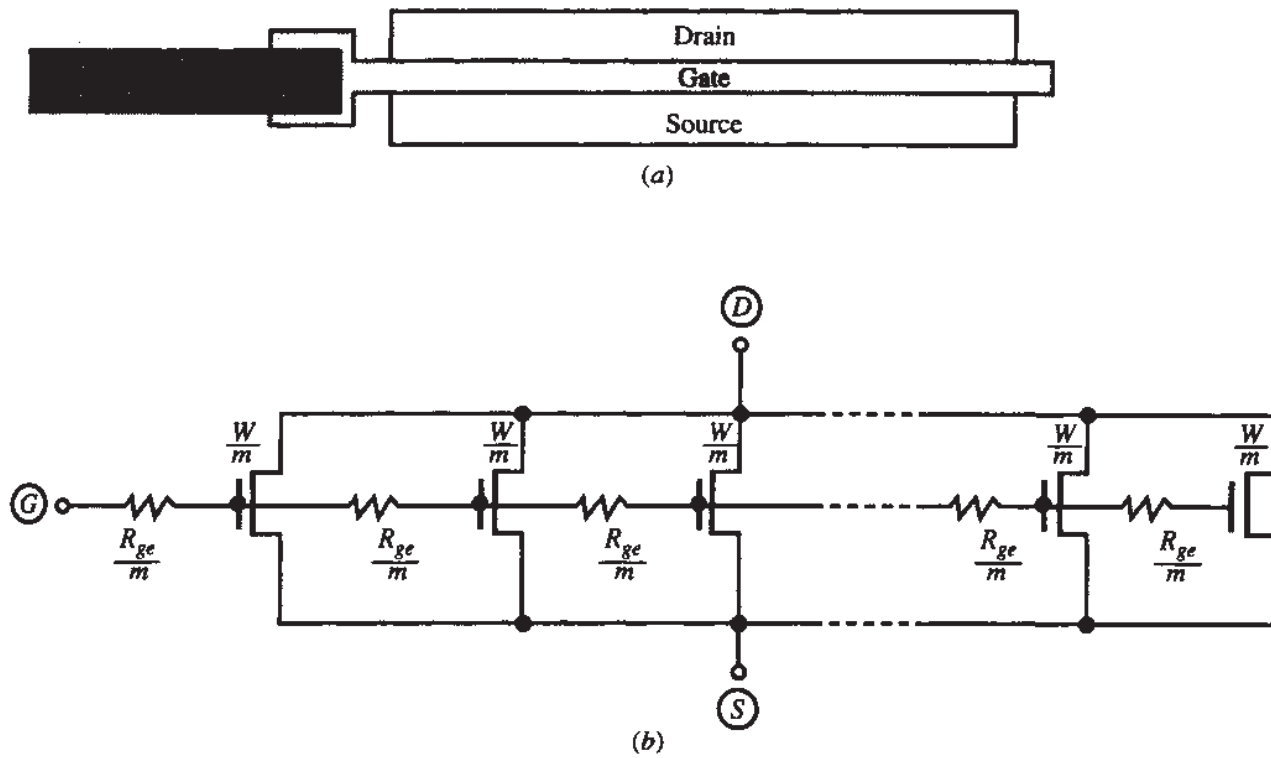
At this point, one may reasonably wonder, “How can such highly simplified models be of any use at RF?” There are several answers to this question. One is related to the extrinsic parasitics; these can dominate device behavior in some cases, and thus limit its application to frequencies lower than those at which the effect of R_{gs} or of τ_1 would be felt. Another answer is related to parameter extraction. The elements are often assigned whatever values make the model give results close to measurements. The values one ends up with in this way are nonphysical, but are used to make up for the inadequacies of the model. This is a dangerous practice, as those

same nonphysical values may result in a very wrong prediction of a different performance parameter, not considered during parameter extraction; this problem will be discussed in more detail in Chap. 10. In general, the fact that a grossly simplified model happens to give satisfactory results for some cases is no reason for trusting such a model; there can be other cases where the same model will prove very inadequate. Simplified models can totally fail to predict certain phenomena which more complete models such as the one in Fig. 9.27a would predict. For example, at high frequencies the impedances of C_{bse} and C_{bde} in that model become very small, and they tend to shunt the channel with the substrate resistances R_{be1} and R_{be3} . This can affect the value of the admittance y_{dd} seen at the drain terminal. This effect cannot be predicted by the model of Fig. 9.28b. In general, models developed for low-frequency work will have difficulty in modeling RF operation for all situations. If such models have to be used, one may at least extend their range of validity by using the approach suggested in Fig. 7.12. This has been attempted for RF applications.⁸¹

In the derivations for the various parameters in this chapter, short-channel effects were not considered. The topologies for small-signal non-quasi-static equivalent circuits derived in this chapter are still useful in the presence of such effects, but the values of the parameters in them must be modified. The influence of some short-channel effects has been considered, to some extent, in the literature.^{49,50,52} Simplified models like the one in Fig. 9.28b are still used in association with short-channel devices, with element values determined from measurements.

It should be noted that, rather than try to stretch simplified models like the one in Fig. 9.28b to fit measurements, one can in principle use a general y -parameter representation, such as the ones shown in Figs. 9.15 to 9.17, since these were derived for an arbitrary four-terminal structure without *any* assumptions as to channel length, substrate uniformity, type of extrinsic effects, etc. In other words, those representations were completely general, and can in principle take care of all distributed effects, at any frequency, provided appropriate values for the admittances in them can be provided as a function of frequency. However, calculating such values is very involved even for long-channel devices.⁴⁵ If values directly extracted from measurements are used, the models do not have any predictive power for situations other than those measured. In addition, y -parameter models are not supported in many simulation programs.

GATE RESISTANCE. In many RF applications, devices with very large gate widths W are used. In such cases, gate resistance effects can be very significant. Consider the device shown in Fig. 9.29a. It is clear that the gate resistance forms a distributed RC circuit with the gate-channel capacitance. To study this effect,⁸²⁻⁸⁶ one can split the device into m subdevices, obtaining the lumped approximation shown in Fig. 9.29b. Here R_{ge} represents the total gate resistance, which in Fig. 9.29a is given by $(W/L)R_{\square}$, where R_{\square} is the sheet resistance of the gate (see Sec. 1.3.2). It is clear that the gate resistance is of little consequence for the subdevices toward the left, but becomes increasingly important for those toward the right; the gate signal those subdevices see will be significantly phase-shifted, due to the resistances and gate capacitances on their left. Writing basic equations, and letting m approach infinity, it can be

**FIGURE 9.29**

(a) A simple transistor layout. (b) Lumped approximation of (a).

shown^{75,85} that the distributed effect of the gate resistance can be approximated by using a single transistor, with an effective gate resistance of

$$R_{ge,eff} = \frac{1}{3} \frac{W}{L} R_{\square} \quad (9.6.1)$$

This value can be used in lieu of R_{ge} in small-signal models like those in Figs. 9.27b and 9.28b.

The gate resistance, in addition to affecting the frequency response, will also contribute noise. It can be shown that, at low frequencies, this noise can again be modeled by using a single effective value for the gate resistance, which is again given by (9.6.1).⁸⁵ Depending on geometrical dimensions and gate resistivity, this noise can in some cases be significantly higher than that produced by the intrinsic part of the device. At high frequencies, the gate resistance noise tends to be “filtered out” by the gate capacitance, and the total noise approaches that produced by the intrinsic part.⁸⁶

If the gate is contacted on both sides, as in Fig. 9.30, the effect will be equivalent to having two devices, each with gate width $W/2$ and thus gate resistance of half the total value, in parallel; thus, in comparison to Fig. 9.28a, the effective resistance will be 4 times smaller, and

$$R_{ge,eff} = \frac{1}{12} \frac{W}{L} R_{\square}, \quad \text{contacts on both sides} \quad (9.6.2)$$

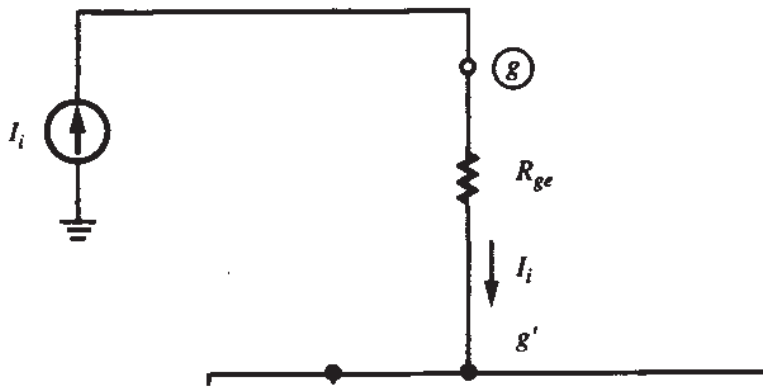


FIGURE 9.30
A transistor with gate contacts on both sides.

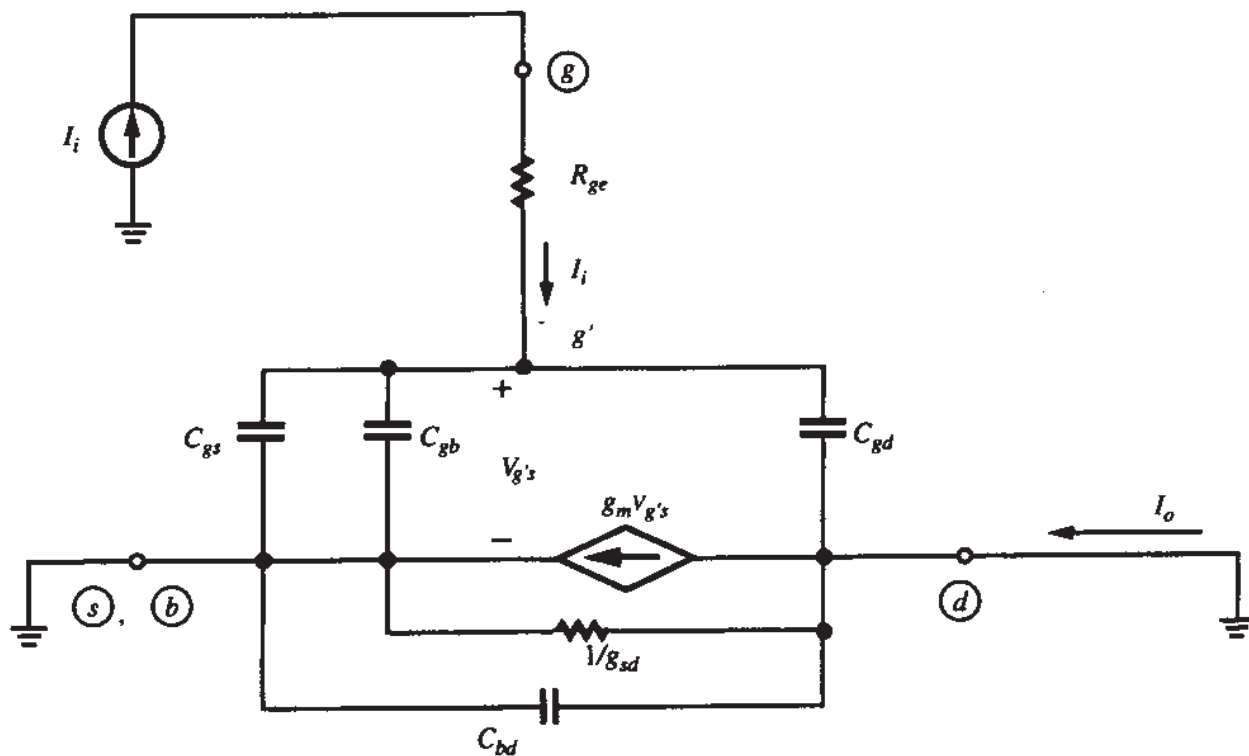


FIGURE 9.31
Circuit for estimating the transition frequency ω_T .

TRANSITION FREQUENCY. In Sec. 8.3.2, we considered the intrinsic part of the transistor and we derived an expression for the “intrinsic transition frequency.” For the complete device, including extrinsic elements, we can similarly calculate a *transition frequency* (or *cutoff frequency*), as that frequency at which the “short-circuit current gain” of the transistor (measured in saturation, with a dc source between drain and source, as in Fig. 8.20a) becomes unity. An estimate can be obtained by using a greatly simplified equivalent circuit, as shown in Fig. 9.31; the short between drain and source corresponds to the dc source in the actual circuit. Voltage and current pha-

sors are indicated in the figure. A current source input is assumed for convenience. The total capacitance between the internal node g' and ground is seen to be

$$C_g = C_{gs} + C_{gb} + C_{gd} \quad (9.6.3)$$

where the individual capacitances include both intrinsic *and* extrinsic components. Thus $V_{g's}$ in the figure is $I_i/(j\omega C_g)$. Elements $1/g_{sd}$ and C_{bd} have a zero voltage across them, and their current is zero; neglecting the small current through C_{gd} , the output current is $I_o = g_m V_{g's} = g_m I_i/(j\omega C_g)$, and thus the short-circuit current gain is $I_o/I_i = g_m/(j\omega C_g)$. The transition frequency ω_T is defined as the frequency at which the magnitude of I_o/I_i becomes unity. Thus

$$\boxed{\omega_T = \frac{g_m}{C_g}} \quad (9.6.4)$$

Let us consider two special cases of interest. Assume first that no velocity saturation is present. Then, g_m can be calculated from (8.2.17a). Using values given for the various capacitances in Secs. 8.3 and 8.4, it can be seen that a simple estimate for C_g is the total oxide capacitance, $C'_{ox}WL$. Using these expressions in (9.6.4) we obtain

$$\boxed{\omega_T = \frac{\mu(V_{GS} - V_T)}{\alpha L^2} = \omega_o, \quad \text{no velocity saturation}} \quad (9.6.5)$$

where ω_o is the quantity defined in (9.4.67). If, instead, complete velocity saturation is assumed, g_m can be calculated from (8.2.18b); from this, and using the same estimate as above for C_g , we obtain

$$\boxed{\omega_T = \frac{|v_d|_{\max}}{L}, \quad \text{velocity saturation}} \quad (9.6.6)$$

where $|v_d|_{\max}$ is the magnitude of the saturated carrier velocity (Sec. 6.5).

Example. For a transistor with a 0.25- μm gate length, assuming a saturation velocity of 10^7 cm/s, the above equation gives $\omega_T = 400$ Grad/s, corresponding to $f_T = \omega_T/(2\pi) = 64$ GHz.

Several observations can now be made. Consider first the effect of reducing L on ω_T . As seen from the above equations, for long-channel devices, reducing the channel length increases ω_T drastically, due to the presence of L^2 in (9.6.5). When the channel becomes very short, though, velocity saturation can occur and thus the increase in ω_T will be less drastic, as seen from (9.6.6). The effect of the gate voltage is also noteworthy. In the absence of velocity saturation effects, ω_T can be increased by

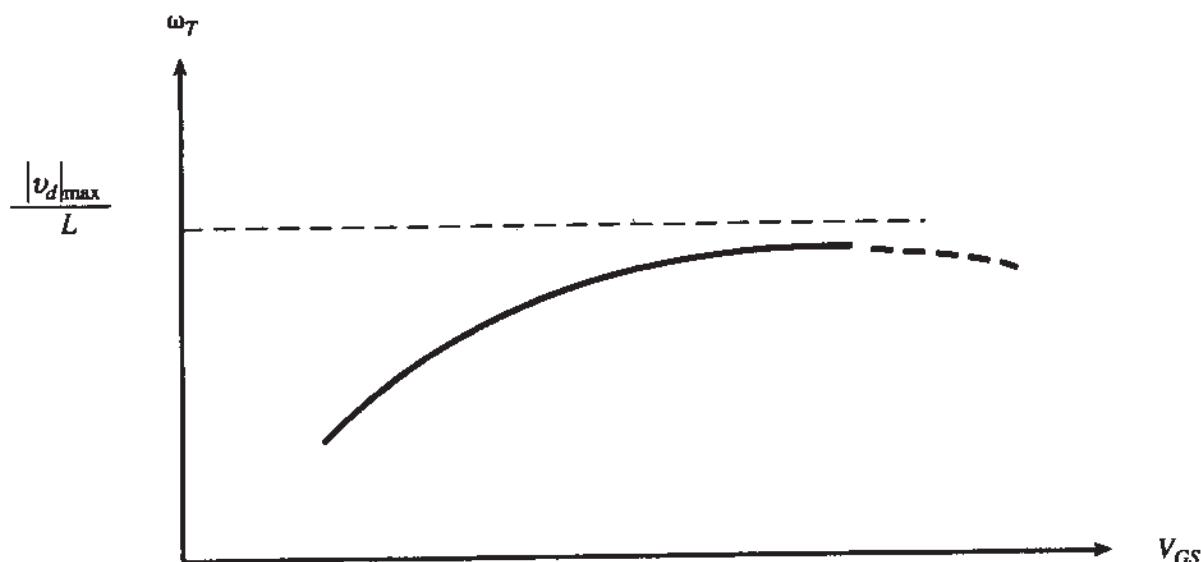


FIGURE 9.32

Transition frequency versus gate-source voltage.

increasing V_{GS} , as seen in (9.6.5). The increase will not be linear, though, since increasing V_{GS} decreases the effective mobility. For large V_{GS} , the saturation voltage V'_{DS} can be large enough for velocity saturation to set in, and thus the effect of V_{GS} diminishes, as can be expected from (9.6.6). This is illustrated in Fig. 9.32. In fact, if the influence of V_{GS} on mobility and of all the various short-channel effects on g_m and C_g in (9.6.4) are considered, it is often found that further increases in V_{GS} can actually cause a decrease in ω_T , as shown toward the right in Fig. 9.32.

More accurate models give more complicated expressions for ω_T , but the corresponding values are usually not too different from what (9.6.4) gives. Thus, this equation is widely used. Notice that R_{ge} does not enter into this expression, since R_{ge} in Fig. 9.31 is in series with a current source, and thus the current entering node g' is I_i , independent of the value of R_{ge} .[†] However, R_{ge} is known to have a detrimental effect in RF circuits, so its absence from the expression for ω_T simply points to the fact that ω_T is not an adequate figure of merit for general RF performance. An additional figure of merit, which makes the effect of R_{ge} clear, is now discussed.

MAXIMUM FREQUENCY OF OSCILLATION. A common figure of merit for RF work is the *maximum frequency of oscillation* ω_{max} , also called *unity power gain frequency*. Power gain in a circuit is the ratio of the load power to the input power. Transistors take power from the power supply, and convert it to output power delivered to a load and controlled by the input. A power gain larger than unity is what qualifies these devices as “active.” Power gain can be maximized by choosing the circuitry connected

[†]More complete models (e.g., the one in Fig. 9.27a) would predict some dependence of ω_T on R_{ge} .

to the transistor in an optimum way: The transistor input is conjugate-matched to the input signal source, the load is conjugate-matched to the transistor output impedance, and an appropriate network is used to cancel the effect of feedback from the output to the input.⁸⁷ The power gain obtained in this way is called *unilateral* power gain. As the frequency of operation is raised, the unilateral power gain decreases and eventually becomes less than unity; the device, then, behaves as a “passive” device. The frequency at which the unilateral power gain drops to unity, is defined as ω_{\max} , and is a quantity of fundamental importance. This quantity is called the *maximum oscillation frequency*, since this is the maximum frequency at which the device can be made to oscillate by feeding its output back to its input. Since R_{ge} dissipates input power, the quantity ω_{\max} captures the effect of this element.

Using circuit theory, one can calculate ω_{\max} for a given model.⁸⁸ For complete models, the resulting expressions can be very complicated, so usually simple estimates are obtained by using simple models. For example, assuming that the source extrinsic resistance R_{se} is very small, one can use the model in Fig. 9.31 (without the output short circuit, and embedded in appropriate circuitry). For this model it can be found that using optimum external circuitry^{87,88} the unilateral power gain is approximately $\omega_T^2/[4R_{ge}\omega^2(g_{sd} + \omega_T C_{gd})]$. The value of ω_{\max} can be found by setting this gain equal to 1, and solving for $\omega = \omega_{\max}$; a consideration of distributed effects⁸⁵ shows that, to take such effects into account, R_{ge} in the resulting expression should be replaced by $R_{ge,\text{eff}}$, as calculated above. Thus we find

$$\omega_{\max} \approx \frac{\omega_T}{\sqrt{4R_{ge,\text{eff}}(g_{sd} + \omega_T C_{gd})}}, \quad R_{se} \ll R_{ge} \quad (9.6.7)$$

This can be larger or smaller than ω_T . The value of the gate resistance is of obvious importance here; unless it is small, it can make ω_{\max} significantly less than ω_T , thus limiting the usefulness of the device for many applications. Consider the effect of increasing W in Fig. 9.29a on ω_T and ω_{\max} . In (9.6.4), both numerator and denominator are proportional to W , so ω_T as calculated from that expression is independent of W . Thus, the numerator of the expression for ω_{\max} above is independent of W ; however, in the denominator both factors are proportional to W [see (9.6.1) and Sec. 8.3]. Thus ω_{\max} is, to first order, inversely proportional to W in this device.

Example. Consider the device of the previous example, and assume that it has $R_{ge,\text{eff}} = 40 \, \Omega$, $g_{sd} = 2 \, \text{mA/V}$, and $C_{gd} = 3 \, \text{fF}$. Then (9.6.7) gives $\omega_{\max} = 559 \, \text{Grad/s}$, or $f_{\max} = \omega_{\max}/(2\pi) = 89 \, \text{GHz}$.

To keep $R_{ge,\text{eff}}$ small one needs silicided gates (Sec. 6.8), multiple contacts (Fig. 9.30), or a device split into several subdevices connected in parallel; while devising appropriate layouts for this, though, care has to be taken not to increase the parasitic capacitance values significantly.

Because of the decisive effect of R_{ge} on ω_{\max} , the inclusion of only this extrinsic resistance in the above calculation is justified in many cases. However, if R_{ge} is made small using the techniques just mentioned, ω_{\max} increases to the point where the

effect of other elements (e.g., of the source series resistance R_{se} , or of R_{gs}) is felt. In such cases, more complete models may have to be used to calculate ω_{\max} . We note that in the literature on RF devices, sometimes the values of ω_{\max} quoted are calculated from a simplified version of (9.6.7), in which either C_{gd} or g_{sd} are assumed to be negligible. This can give artificially high ω_{\max} values in some cases.

Calculating ω_T and ω_{\max} from more complete models than the ones used above leads to very complicated expressions. Thus, the simple expressions we have presented are widely used. The quantities ω_T and ω_{\max} are often measured by extrapolation from the transistor behavior at low frequencies; thus, the expressions shown above, which were derived using low-frequency models, are consistent with this practice.

REFERENCES

1. J. A. Geurst, "Calculation of high-frequency characteristics of thin film transistors," *Solid-State Electronics*, vol. 8, pp. 88–90, 1965.
2. D. B. Candler and A. G. Jordan, "A small-signal, high-frequency analysis of the insulated-gate field-effect transistor," *International Journal of Electronics*, vol. 19, pp. 181–196, 1965.
3. M. H. White and R. C. Gallagher, "Metal oxide semiconductor (MOS) small-signal equivalent circuits," *Proceedings of the IEEE*, vol. 53, pp. 314–315, 1965.
4. J. R. Hauser, "Small-signal properties of field-effect devices," *IEEE Transactions on Electron Devices*, vol. ED-12, pp. 605–618, 1965.
5. J. A. Geurst and H. J. C. A. Nunnink, "Numerical data on the high-frequency characteristics of thin-film transistors," *Solid-State Electronics*, vol. 8, pp. 769–771, 1965.
6. W. Fischer, "Equivalent circuit and gain of MOS field-effect transistors," *Solid-State Electronics*, vol. 9, pp. 71–81, 1966.
7. R. Paul, "Hochfrequenzverhalten von Feldeffekttransistoren mit isolierter Steuer Elektrode," *AEU*, vol. 20, pp. 317–328, 1966.
8. R. Paul, "Die Ersatzschaltung von Feldeffekttransistoren mit isoliertem Gate," *Nachrichtentechnik*, vol. 16, pp. 243–249, 1966.
9. R. Paul, "Einfluss einer nichtidealen Gateisolation auf die Vierpolparameter des Feldeffekttransistors," *Nachrichtentechnik*, vol. 16, pp. 278–285, 1966.
10. R. Paul, "Frequenzabhängigkeit der Vierpoleigenschaften von MOS-Transistoren," *Nachrichtentechnik*, vol. 16, pp. 401–406, 1966.
11. Z. S. Girbnikov and Yu. A. Tkhorik, "Calculation of the transient processes in field-effect triodes with an insulated gate for the saturated mode of operation," *Radio Engineering and Electronic Physics*, vol. 11, pp. 776–781, 1966.
12. H. Johnson, "A high-frequency representation of the MOS transistor," *Proceedings of the IEEE*, vol. 54, pp. 1970–1971, 1966.
13. D. H. Treleaven and F. N. Trofimenkoff, "MOSFET equivalent circuit at pinchoff," *Proceedings of the IEEE*, vol. 54, pp. 1223–1224, 1966.
14. H. C. DeGraaff, "High frequency measurements of thin-film transistors," *Solid-State Electronics*, vol. 10, pp. 51–56, January 1967.
15. R. Paul, "Experimentelles Hochfrequenzverhalten von MOS-Transistoren," *Nachrichtentechnik*, vol. 17, pp. 255–260, July 1967.
16. M. B. Das, "Generalized high-frequency network theory of field-effect transistors," *IEE Proceedings*, vol. 114, pp. 50–59, 1967.
17. I. R. Burns, "High-frequency characteristics of the insulated gate field-effect transistors," *RCA Review*, vol. 28, pp. 385–418, 1967.

18. F. A. Lindholm, R. J. Balda, and J. L. Clements, "Characterization of the four-terminal MOS transistor for digital and linear applications," *Digest of Technical Papers*, International Electronics Conference, Toronto, pp. 116–117, 1967.
19. M. B. Das, "High-frequency network properties of MOS transistors including the substrate resistivity effects," *IEEE Transactions on Electron Devices*, vol. ED-16, pp. 1049–1069, 1969.
20. J. A. Van Nielen, "A simple and accurate approximation to the high-frequency characteristics of IGFETs," *Solid-State Electronics*, vol. 12, pp. 826–829, 1969.
21. J. W. Haslett and F. N. Trofimenkoff, "Small-signal, high-frequency equivalent circuit for the metal-oxide semiconductor field-effect transistor," *IEE Proceedings*, vol. 116, pp. 699–702, 1969.
22. R. S. C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley-Interscience, New York, 1970.
23. E. M. Cherry, "Small-signal high-frequency response of the insulated gate field-effect transistor," *IEEE Transactions on Electron Devices*, vol. ED-17, pp. 569–577, 1970.
24. M. V. Balakirev and V. M. Bogachev, "Frequency and transient characteristics of metal oxide semiconductor transistors and synthesis of their equivalent circuits," *Radio Engineering and Electronics Physics*, vol. 16, pp. 1884–1897, 1971.
25. M. Reiser, "A two-dimensional numerical FET model for DC, AC and Large-Signal Analysis," *IEEE Transactions on Electron Devices*, vol. ED-20, pp. 35–45, 1976.
26. U. Kumar and S. C. Dutta Roy, "A simple small-signal two-part MOST model for the pre-pinchoff region," *Solid-State Electronics*, vol. 20, pp. 1021–1022, 1977.
27. U. Kumar, "A simple two-part model of the metal oxide semiconductor transistor," *Microelectronics Journal*, vol. 10, pp. 50–53, 1978.
28. J. I. Arreola, "Equivalent circuit modeling of the large signal transient response of four-terminal MOS field-effect transistors," doctoral dissertation, University of Florida, 1978.
29. D. E. Ward and R. W. Dutton, "A charge-oriented model for MOS transistor capacitances," *IEEE Journal of Solid-State Circuits*, vol. SC-13, pp. 703–707, October 1978.
30. J. A. Robinson, Y. A. El-Mansy, and A. R. Boothroyd, "A general four-terminal charging-current model for the insulated-gate field effect transistor," *Solid-State Electronics*, vol. 23, parts I, II, pp. 405–414, 1980.
31. S. Y. Oh, "A simplified two-dimensional numerical analysis of MOS devices including transient phenomena," Technical Report G201-10, Integrated Circuits Laboratory, Stanford University, California, June 1981.
32. D. E. Ward, "Charge-based modeling of capacitance in MOS transistors," Technical Report G201-11, Integrated Circuits Laboratory, Stanford University, California, June 1981.
33. R. Conilogue and C. Viswanathan, "A complete large and small signal charge model for a MOS transistor," *Technical Digest*, International Electron Devices meeting, pp. 654–657, San Francisco, 1982.
34. K. Y. Tong, "AC model for MOS transistors from transient-current computations," *IEE Proceedings*, vol. 130, part I, pp. 33–36, February 1983.
35. J. J. Paulos and D. A. Antoniadis, "Limitations of quasi-static capacitance models for the MOS transistor," *IEEE Electron Device Letters*, vol. EDL-4, pp. 221–224, July 1983.
36. C. Turchetti, G. Masetti, and Y. Tsividis, "On the small-signal behavior of the MOS transistor in quasi-static operation," *Solid-State Electronics*, vol. 26, pp. 941–949, 1983.
37. Y. Tsividis and G. Masetti, "Problems in precision modeling of the MOS transistor for analog applications," *IEEE Transactions on Computer-Aided Design*, vol. CAD-3, pp. 72–79, January 1984.
38. J. J. Paulos, "Measurement and modeling of small-geometry MOS transistor capacitances," Ph.D. dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, September 1984.
39. M. Bagheri and Y. Tsividis, "A small-signal dc-to-high-frequency nonquasi-static model for the four-terminal MOSFET valid in all regions of operation," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 2383–2391, November 1985.
40. M. Bagheri, "Improving the non-quasi-static weak-to-strong-inversion four-terminal MOSFET model," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 2558–2560, December 1987.

41. K.-C. Chai and J. J. Paulos, "Comparison of quasi-static and non-quasi-static capacitance models for the four-terminal MOSFET," *IEEE Electron Device Letters*, vol. EDL-8, pp. 377-379, September 1987.
42. U. Kumar, "Modified charge-control model for MOS transistors in pre-saturation region," *Solid-State Electronics*, vol. 31, pp. 999-1001, 1988.
43. K.-W. Chai and J. J. Paulos, "Unified nonquasi-static modeling of the long-channel four-terminal MOSFET for large- and small-signal analyses in all operating regimes," *IEEE Transactions on Electron Devices*, vol. 36, pp. 2513-2520, November 1989.
44. P. J. V. Vandeloo and W. M. C. Sansen, "Modeling of the MOS transistor for high frequency analog design," *IEEE Transactions on Computer-Aided Design*, vol. 8, pp. 713-723, July 1989.
45. L.-J. Pu and Y. Tsividis, "Small-signal parameters and thermal noise of the four-terminal MOSFET in non-quasistatic operation," *Solid-State Electronics*, vol. 33, p. 521, 1990.
46. L.-J. Pu and Y. Tsividis, "Harmonic distortion of the four-terminal MOSFET in non-quasistatic operation," *IEE Proceedings*, vol. 137, pp. 325-332, October 1990.
47. K. A. Sakallah, Y.-T. Yen, and S. S. Greenberg, "A first-order charge conserving MOS capacitance model," *IEEE Transactions on Computer-Aided Design*, vol. 9, pp. 99-108, January 1990.
48. P. Roblin, S. C. Kang, and W.-R. Liou, "Improved small-signal equivalent circuit model and large-signal state equations for the MOSFET/MODFET wave equation," *IEEE Transactions on Electron Devices*, vol. 38, pp. 1706-1718, August 1991.
49. H. J. Park, P. K. Ko, and C. Hu, "A non-quasi-static MOSFET model for SPICE-AC analysis," *IEEE Transactions on Computer-Aided Design*, vol. 11, pp. 1247-1257, October 1992.
50. S. C. Kang and P. Roblin, "Optimal second-order small-signal model for long- and short-channel three-terminal MOSFET/MODFET wave equation," *IEEE Transactions on Electron Devices*, vol. 39, pp. 1909-1915, August 1992.
51. R. Singh, A. Juge, R. Joly, and G. Morin, "An investigation into the nonquasi-static effects in MOS devices with on-wafer s-parameter techniques," *Proceedings of the International Conference on Microelectronic Test Structures*, pp. 21-25, 1993.
52. T. Smedes, and F. M. Klaassen, "An analytical model for the nonquasi-static small-signal behaviour of submicron MOSFETs," *Solid-State Electronics*, vol. 38, pp. 121-130, 1995.
53. J. Katzenelson and A. Unikovski, "A network charge-oriented MOS transistor model," *International Journal of High Speed Electronics and Systems*, vol. 6, pp. 285-316, 1995.
54. V. I. Kol'dyaev, A. Clerix, and L. Deferm, "Closed-form frequency dependent gate-to-channel capacitance model for submicron MOSFET's," *Proceedings of the 36th European Solid-State Device Research Conference*, pp. 679-682, Bologna, Italy, September 1996.
55. Y. Niitsu, "Simple small-signal model for 3-port MOS transistors," *IEICE Transactions on Electronics*, vol. E79-C, pp. 1760-1765, December 1996.
56. J. J. Paulos, D. A. Antoniadis, and Y. P. Tsividis, "Measurement of intrinsic capacitances of MOS transistors," *Digest of Technical Papers, IEEE International Solid-State Circuits Conference*, San Francisco, pp. 238-239, February 1982.
57. C. A. Desoer and E. S. Kuh, *Basic Circuit Theory*, McGraw-Hill, New York, 1969.
58. W. H. Hayt, Jr., and J. E. Kemmerly, *Engineering Circuit Analysis*, McGraw-Hill, New York, 1993.
59. W. H. Kim and H. E. Meadows, Jr., *Modern Network Analysis*, John Wiley, New York, 1971.
60. M. Shoji, "Analysis of high-frequency thermal noise of enhancement mode MOS field-effect transistors," *IEEE Transactions on Electron Devices*, vol. ED-13, pp. 520-524, June 1996.
61. H. E. Halladay and A. Van Der Ziel, "On the high frequency excess noise and equivalent circuit representation of the MOS-FET with n -type channel," *Solid-State Electronics*, vol. 12, pp. 161-176, 1969.
62. F. M. Klaassen, "A combination of the high-frequency noise quantities of a MOS-FET," *Philips Research Reports*, vol. 24, pp. 559-571, 1969.
63. F. M. Klaassen and J. Prins, "Noise of field-effect transistors at very high frequencies," *IEEE Transactions on Electron Devices*, vol. ED-16, pp. 952-957, November 1969.
64. P. S. Rao and A. Van Der Ziel, "Noise and y -parameters in MOS FET'S," *Solid-State Electronics*, vol. 14, pp. 939-944, 1971.

65. E. W. Kirk and A. Van Der Ziel, "Induced gate noise in MOS FET'S," *Solid-State Electronics*, vol. 14, pp. 945-948, 1971.
66. A. Ambrozy, *Electronic Noise*, McGraw-Hill, New York, 1982.
67. A. van der Ziel, *Noise in Solid-State Devices and Circuits*, Wiley, New York, 1986.
68. H. C. de Graaff and F. M. Klaassen, *Compact Transistor Modelling for Circuit Design*, Springer-Verlag, Vienna, 1990.
69. D. P. Triantis, A. N. Birbas, and S. E. Plevridis, "Induced gate noise MOSFETs revisited: The sub-micron case," *Solid-State Electronics*, vol. 41, pp. 1937-1942, 1997.
70. C. Raynaud, J. Gautier, G. Guegan, M. Lerme, E. Playez, and G. Dambrine, "High-frequency performance of submicrometer channel-length silicon MOSFET's," *IEEE Electron Device Letters*, vol. 12, pp. 667-669, December 1991.
71. A. E. Schmitz, R. H. Walden, L. E. Larson, S. E. Rosenbaum, R. A. Metzger, J. R. Behnke, and P. A. Macdonald, "A deep-submicrometer microwave/digital CMOS/SOS technology," *IEEE Electron Devices Letters*, vol. 12, pp. 16-17, January 1991.
72. N. Camilleri, J. Costa, D. Lovelace, and D. Ngo, "Silicon MOSFETs, the microwave device technology for the 90s," *IEEE MTT-S Digest*, pp. 545-548, 1993.
73. J. C. Costa, D. Lovelace, D. Ngo, and N. Camilleri, "Modeling a new generation of RF devices: MOSFETs for L-band applications," *IEEE MTT-S Digest*, pp. 293-296, 1993.
74. D. Lovelace, J. Costa, and N. Camilleri, "Extracting small-signal model parameters of silicon MOSFET transistors," *IEEE-MTT-S Digest*, pp. 865-868, 1994.
75. R. R. J. Vanoppen, J. A. M. Geelen, and D. B. M. Klaassen, "The high-frequency analogue performance of MOSFETs," *Digest, International Electron Devices Meeting*, pp. 173-176, December 1994.
76. J. M. Collantes, J. J. Raoux, J. P. Villotte, R. Quere, G. Montoriol, and F. Dupis, "A new large-signal model based on pulse measurement techniques for RF power MOSFET," *IEEE MTT-S Digest*, pp. 1553-1556, 1995.
77. D. B. M. Klaassen, B. Nauta, and R. R. J. Vanoppen, "RF modelling of MOSFETs," in *Analog Circuit Design: MOST RF Circuits, Sigma-Delta Converters, and Translinear Circuits*, W. Sansen, J. H. Huijsing and R. J. v. d. Plassche (editors), Kluwer Academic Publishers, Boston, 1996.
78. S. Lee, H. K. Yu, C. S. Kim, J. G. Koo, and K. S. Nam, "A novel approach to extracting small-signal model parameters of silicon MOSFET's," *IEEE Microwave and Guided Wave Letters*, pp. 75-77, March 1997.
79. M. C. Ho, K. Green, R. Culbertson, J. Y. Yang, D. Ladwig, and P. Ehnis, "A physical large signal Si MOSFET model for RF circuit design," *IEEE MTT-S Digest*, pp. 391-394, 1997.
80. M. Miller, T. Dinh, and E. Shumate, "A new empirical large signal model for silicon RF LDMOS FETs," *IEEE*, pp. 19-22, 1997.
81. M. C. Ho, F. Brauchler, and J. Y. Yang, "Scalable RF Si MOS-FET distributed lumped element model based on BSIM3v3," *Electronics Letters*, vol. 33, pp. 1992-1993, November 1997.
82. H. C. Lin, Y. F. Arzoumanian, J. L. Halsor, M. N. Giuliano, and H. F. Benz, "Effect of silicon gate resistance on the frequency response of MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 255-264, May 1975.
83. T. Sakurai and T. Iizuka, "Gate electrode RC delay effects in VLSI's," *IEEE Journal of Solid-State Circuits*, vol. SC-20, pp. 290-294, February 1985.
84. L.-S. Kim and R. W. Dutton, "Modeling of the distributed gate RC effect in MOSFET's," *IEEE Transactions on Computer-Aided Design*, vol. 8, pp. 1365-1367, December 1989.
85. B. Razavi, R.-H. Yan, and K. F. Lee, "Impact of distributed gate resistance on the performance of MOS devices," *IEEE Transactions on Circuits and Systems—I*, vol. 41, pp. 750-754, November 1994.
86. E. Abou-Allan and T. Manku, "A small-signal MOSFET model for radio frequency IC applications," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 16, pp. 437-447, May 1997.
87. I. Bahl and P. Bhartia, *Microwave Solid-State Circuit Design*, Wiley, New York, 1988.
88. E. S. Kuh and R. A. Rohrer, *Theory of Linear Active Networks*, Holden-Day, San Francisco, 1967.

PROBLEMS

- 9.1. For the case of Fig. P9.1 (where $q = -q_1$) relate the various quantities by writing equations analogous to (7.3.16) and (9.2.2). Show that in this case the signs used in the definition (9.2.1) represent a “natural” choice.

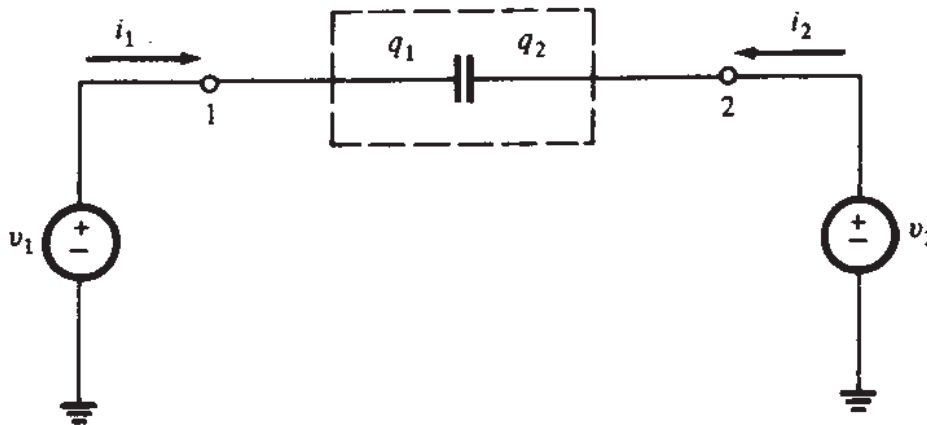


FIGURE P9.1

- 9.2. We define the “nonreciprocity” between two capacitance parameters C_{kl} and C_{lk} ($l \neq k$) as the quantity $C_{kl} - C_{lk}$. Using (9.2.8), derive relations among the various nonreciprocities and comment on the result.
- 9.3. Using Kirchhoff’s current law for all terminals, check the validity of the model in Fig. 9.4b.
- 9.4. Prove in detail that (9.2.19) and (9.2.20) are equivalent to (9.2.12).
- 9.5. Consider the model of Fig. 9.5. For *each* of 16 ordered pairs of terminals (k, l) perform the following experiment. Apply a voltage v_l between terminal l and ground, keeping all other terminals grounded, and determine the current i_k entering terminal k . Show that the capacitive part of this current is given by $C_{kl}(dv_l/dt)$ if $l = k$, and by $-C_{kl}(dv_l/dt)$ if $l \neq k$. Note that six of the C_{kl} appear directly in the model, whereas others do not, in which case (9.2.8) or (9.2.20) will have to be evoked.
- 9.6. Show that results identical to those in Prob. 9.5 would be obtained if the model of Fig. 8.17 were used instead for five of the terminal pair combinations (those that correspond to the subscripts of the capacitors in Fig. 8.17). Show that for other terminal pairs the results will differ from those in Prob. 9.5; this points to the incompleteness of the simple model of Fig. 8.17. Explain why this incompleteness does not matter at low frequencies.
- 9.7. Prove (9.2.21) to (9.2.27).
- 9.8. An intuitive argument was given following (9.2.24) concerning the fact that C_{sd} is negative in nonsaturation. Using similar arguments, explain why C_{ds} is also negative, whereas all other capacitances C_{kl} are positive.
- 9.9. Using the simplified strong-inversion model, $C_{mx} = 0$ was found. Allow for a nonzero C_{mx} , and show that the expressions for the capacitances would have to be modified so that (9.2.8) remains valid.

- 9.10.** This problem continues the comparison between the simple model of Fig. 8.17 and the complete quasi-static model of Fig. 9.5, started in Prob. 9.6. Consider operation in the saturation region. Using the small-signal quantity values found for the simple model in Secs. 8.2 and 8.3, justify the upper frequency limit of validity given for the model of Fig. 8.17 following (8.3.6).
- 9.11.** Prove (9.3.7).
- 9.12.** Prove (9.3.9).
- 9.13.** Provide all the details in the proofs of (9.4.36) to (9.4.48).
- 9.14.** Prove (9.4.51) and (9.4.53) in detail.
- 9.15.** (*Note:* This problem requires *extremely lengthy* derivations. It is not meant as a home-work problem.) Solve (9.4.51) and (9.4.52) using the following iterative procedure.²⁰ Replace x in (9.4.51) with a dummy variable \hat{x} . Develop an integral form of (9.4.51a) and (9.4.51b), by integrating from x to L . Use as an initial approximation $I_i(x, \omega) = I_i(L, \omega)$ (which is actually exact for $\omega = 0$). Substitute this in the integral form of (9.4.51a) and solve for the corresponding approximation of $U_i(x, \omega)$. Use this in the integral form of (9.4.51b) to find a new approximation for $I_i(x, \omega)$. Use this in the integral form of (9.4.51a) and solve for the new approximation of $U_i(x, \omega)$, etc. Show that the procedure produces a solution from which one can obtain (9.4.58) to (9.4.60), with the values given in Appendix N.
- 9.16.** Prove (9.4.65) to (9.4.67), using results from Appendix N. (*Note:* Very lengthy derivations are involved.)
- 9.17.** (a) Plot the magnitude and phase of $-y_{gb}$ versus ω (for ω up to ω_o on a log axis) in the saturation region and at $V_{DS} = 0$, using (9.4.69e) with and without including y_a . Use an oxide thickness of 50 Å, $\alpha = 1.1$, $W = L = 2 \mu\text{m}$, $\mu = 600 \text{ cm}^2/(\text{V}\cdot\text{s})$, and $V_{GS} - V_T = 2 \text{ V}$. Comment on the validity of approximating $-y_{gb}$ by $j\omega C_{gb}$ (which is equivalent to deleting the box shown in broken lines in Figs. 9.20, 9.23, and 9.24). Comment on currents that, in practice, are likely to mask the current due to y_{gb} .
- (b) Assume $V_{DS} = 0$. Show instances where a terminal current being observed is only due to y_{gb} (and to extrinsic elements). Explain why such instances are unlikely to be of significance in practice and that, even then, the magnitude of the observed current is very small and likely to be dominated by extrinsic effects.
- 9.18.** Verify the two circuit representations in Fig. 9.19.
- 9.19.** Using the approximation in (9.4.72), show that the model in Fig. 9.20 reduces at low frequencies to the model of Fig. 9.5, in both form and element values (y_{mx} is assumed zero), and that at even lower frequencies it further reduces to the model of Fig. 8.17.
- 9.20.** The intrinsic cutoff frequency of a transistor was defined in Sec. 8.3, and in the strong-inversion saturation region was found to be given by (8.3.31), by using the model of Fig. 8.17. As mentioned following that equation, the model of Fig. 8.17 is not accurate at frequencies as high as $\omega = \omega_T$, and therefore the result (8.3.31) should be checked by using a more accurate model. Show that, by using the model of Fig. 9.23, one obtains practically the same value for ω_T .
- 9.21.** Prove all the values given in the model in Fig. 9.24.
- 9.22.** Provide plots versus ω for normalized magnitude and phase of all parameters in Fig. 9.17 in the saturation region, for the non-quasi-static model of Sec. 9.4.

- 9.23.** Consider a transistor with $N_A = 10^{17} \text{ cm}^{-3}$, $t_{ox} = 50 \text{ \AA}$, $V_{T0} = 0.5 \text{ V}$, $\phi_0 = 0.9 \text{ V}$, $\mu = 600 \text{ cm}^2/(\text{V}\cdot\text{s})$, and $W = L = 2 \text{ }\mu\text{m}$ connected as shown in Fig. 9.1b with $V_B = 0.5 \text{ V}$, $V_S = 1 \text{ V}$, $V_G = 3 \text{ V}$, and $V_D = 1.5 \text{ V}$. Calculate all parameters for the model of Fig. 9.23.
- 9.24.** For the device of Prob. 9.23, find the magnitude and phase of all small-signal terminal currents (Fig. 9.1b) if the small-signal voltages $v_b(t)$, $v_s(t)$, $v_g(t)$, and $v_d(t)$ are of the form $a \cos(\omega t + \phi)$, with magnitudes a of 2, 1, 4, and 3 mV, respectively, and phases of 0, 0, 40, and 50°, respectively. The frequency ω is $2\pi \times 5 \text{ GHz}$.
- 9.25.** The circuit in Fig. P9.25 has little to do with a transistor, but it will help in getting a feeling for the various levels of approximation involved in modeling.
- Find the total admittance y_{AB} , its real part y_{ABr} , and its imaginary part y_{ABi} .
 - Find the conditions under which C can be neglected in each of the above three expressions, and give a simplified model.
 - Find the conditions under which R_2 can be neglected in each of the three expressions in (a), and give a simplified model.
 - Show that neglecting R_2 is equivalent to assuming that the charge Q on the top plate depends quasi-statically on the voltage V defined in the figure. Show that at high frequencies this quasi-static dependence breaks down.
- 9.26.** Show that the circuits of Fig. 9.22b and c both predict the behavior in (9.4.71), and find the proper value to be used for C in order for this behavior to be identical even quantitatively. Attempt to explain the value found for C . Equation (8.3.13b) may be helpful.
- 9.27.** Consider a MOS transistor with its body connected to source, driven by a bias plus small-signal voltage between gate and source, and with its drain not connected to anything (Fig. P9.27a). In this connection $I_D = 0$, $V_{DS} = 0$, and the channel is uniform. Derive an expression for the small-signal voltage gain of this device, V_{ds}/V_{gs} , where V_{ds} and V_{gs} are phasors corresponding to small-signal voltages, assuming that the substrate doping is very light, and thus the substrate parasitics can be neglected; also, assume that W and L are very large, so that intrinsic effects dominate and extrinsic parasitics can be neglected. Show that $|V_{ds}/V_{gs}|$ behaves as in Fig. P9.27b; i.e., it can exceed unity in a range of frequencies. Show that this behavior is also predicted by the simple lumped model of Fig. P9.27c. Are the findings in this problem strange? Why, or why not?

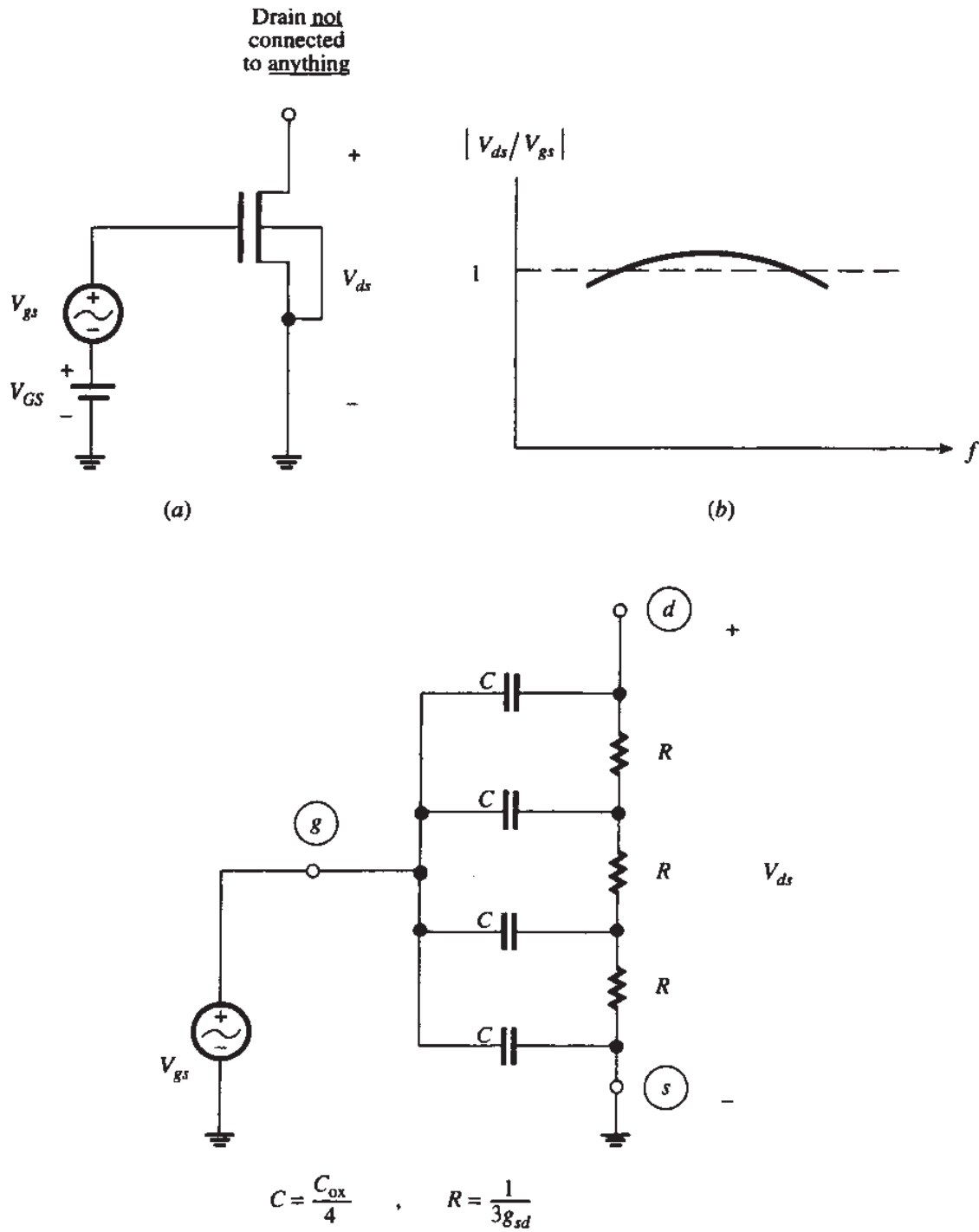


FIGURE P9.27

9.28. Derive (9.5.3) from (9.5.2).

9.29. Shown that Fig. P9.29 is the model of Fig. 9.24 in saturation with three important extrinsic parasitics added, assuming that the substrate is short-circuited to the source and that C_{gb} and L_{sd} are negligible. Assume $R_{de} \ll 1/g_{sd}$ and $\omega \ll 1/(C_{bde}R_{de})$. Derive expressions for the y -parameters y_{gg} , y_{gd} , y_{dd} , and y_{dg} . Show qualitative plots for their real and imaginary parts vs. ω on log-log axes for (a) zero C_{gde} , C_{bde} , and R_{de} , and (b) nonzero C_{gde} , C_{bde} , and R_{de} . Discuss the important effects caused by the extrinsic elements.

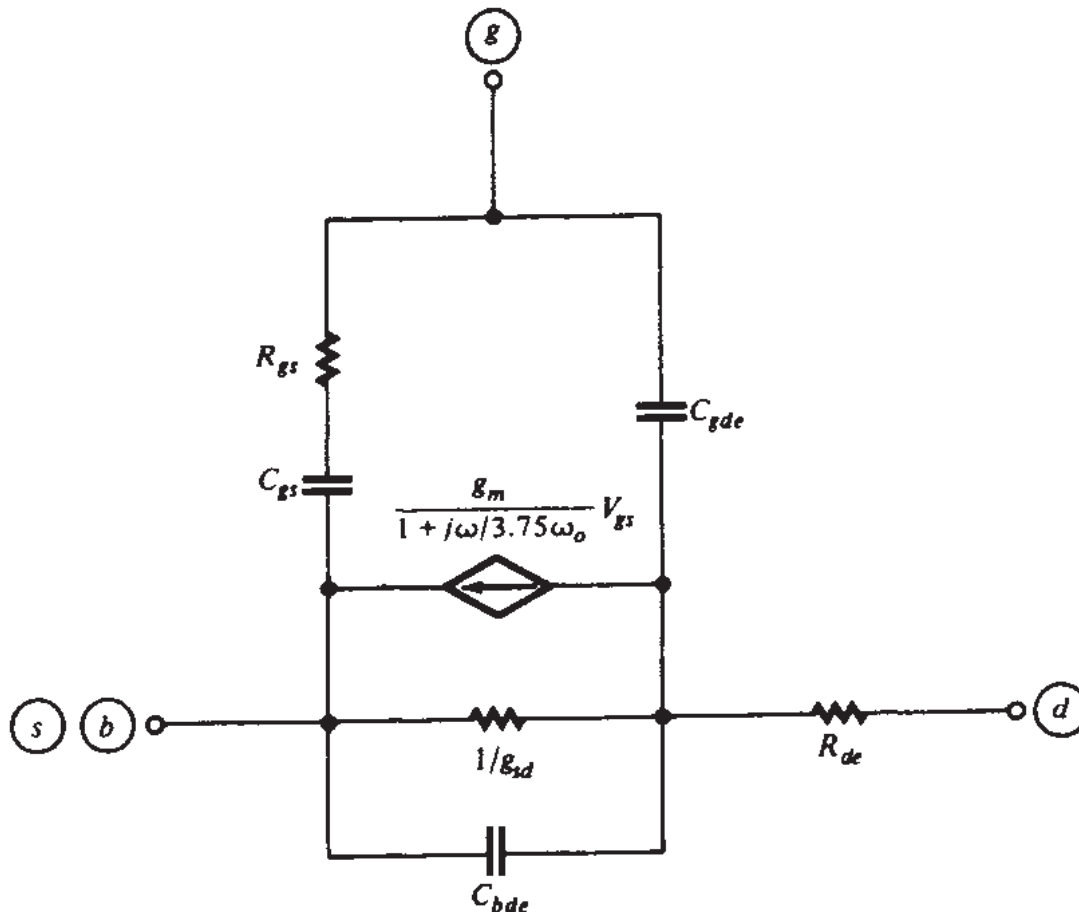


FIGURE P9.29

CHAPTER 10

MOSFET MODELING FOR CIRCUIT SIMULATION

10.1 INTRODUCTION

Computer simulation is today a standard part of integrated circuit design.¹⁻¹⁰ During this process, the computer solves a large set of equations describing (1) the connections between the various circuit elements and (2) the models of these elements. Circuits to be simulated can contain very large numbers of elements. The model for each element can contain a large number of equations. The behavior of the entire circuit may be needed at many points, e.g., 1000 times points for a transient simulation, and all equations may have to be solved repeatedly for each point, as part of numerical iterations for solving the implicit equations. Thus, some computer simulations are very time-consuming. Computing time statistics show that often most of the computation time is spent in evaluating the quantities that are described by the device model equations. It follows, then, that the model equations must be as simple as possible, provided accuracy is not compromised. Over the years, many MOS transistor models for use in computer simulation have been described.¹¹⁻⁸⁴ In this chapter, we will discuss several considerations that are involved in creating such models for today's needs.

A rigorous way of describing the operation of a transistor is to write the fundamental semiconductor equations in three dimensions. These will be coupled nonlinear partial differential equations, one for each of thousands of finite volume elements in the device. Programs are available for setting up such equations and solving them numerically.⁸⁵⁻⁹¹ Although such programs are invaluable for device (as opposed to circuit) analysis and design, the solution can take a long time even for a single

transistor. Such an approach is out of the question for general circuit simulation. Much more efficient models are thus needed, which describe the electrical behavior analytically. These are called *compact* models, or CAD models, and are the subject of this chapter. In what follows, the word *model* will refer exclusively to such compact models, unless it is specifically mentioned otherwise. The word *technology*, or *process*, will refer to a given fabrication process.

10.2 TYPES OF MODELS

There are three main types of compact models for MOS transistors:

1. *Physical models.* These are based on device physics. The parameters in these models have a physical significance (the flat-band voltage, the substrate doping concentration, etc.). Most of the models we have seen in this book are physical (sub)models describing specific phenomena in a transistor.
2. *Empirical models.* A purely empirical model relies on just curve fitting. It can use any equation that adequately fits data (e.g., a polynomial or a cubic spline function). The parameters in an empirical model are just the coefficients, exponents, etc., used in the curve-fitting expressions, and have no physical significance.
3. *Table models.*⁷¹⁻⁸³ These are typically in the form of tables containing the values of the drain current (and, in some cases, of small-signal parameters) for a large number of combinations of bias voltages. The computer then "looks up" such values, instead of calculating them, thus saving time. The values stored can come from measurements, or from a numerical device simulator. A very large number of values must be stored if accuracy is desired, and interpolation functions must be provided for computing the values for points in between those stored. Variations exist, where a smaller number of points are stored, along with the coefficients for extensive interpolation functions.

Empirical and table models provide values independent of which phenomena are responsible for those values. They can thus be developed fast, can be used for a number of technologies, and can be quickly updated. On the other hand, a different set of empirical expression parameters, or a different table, is needed for each combination of W and L values and each temperature value, since these models have no way to incorporate small dimension and temperature effects. These models are best used for dc I - V characteristics, and even for them it is not easy to reliably predict the current for biases outside the range in which the model was optimized, or for W , L , or temperature values different from those used in the fitting process. Finally, these models cannot be used to predict what would happen if the device parameters changed. In particular, they cannot be used for statistical modeling; in other words, it is not possible to provide the statistics describing the systematic and random variations in the fabrication process parameters, such as substrate doping and oxide thickness, and use these models to predict the range of variations of the I - V characteristics.

Table models have found some special-purpose uses, particularly in cases where no good physical models have been developed. Fully empirical models are hardly used per se, but often empirical expressions can be found in table models (to do the interpolation) and in physical models, to help predict some effects whose physics resists a simple analytical description. Save for this last usage, we will not discuss these models here. The rest of this chapter focuses on physical models.

Physical models take a long time to develop (in fact, they have been under development for 30 years, and we still do not have models that are satisfactory for all circuit design uses!). Extra development time may be needed each time a fabrication technology changes significantly; for example, many of the small-dimension phenomena discussed in Chap. 6 were revealed as technology matured, and this necessitated new research and development to make adequate models possible. In addition, if physical models are to be computationally efficient, they must be kept simple, and thus their accuracy must usually be compromised to some degree. Once a good physical model is developed, though, its advantages, in comparison to empirical and table models, are very significant. To begin with, the parameters in a physical model have a physical significance. It is thus possible to relate the results provided by the model to the physical details of the transistor, which is very important in integrated circuit design. Notably, the effects of the geometry of the device (W and L values), and the effects of temperature changes, can be predicted for a given process a designer is working with. Within limits, it is even possible to predict what would happen if the process parameters were changed. This latter feature is of particular significance in statistical modeling. If the basic parameters of a model are the parameters of the process (e.g., substrate doping concentration, oxide thickness), one can use information about the systematic and random errors of these parameters, as well as such errors in the geometrical dimensions, to predict the statistics of the I - V characteristics or other device behavior (e.g., the transistor f_T). Thus, physical models can be used for statistical analysis,⁹²⁻¹⁰¹ in order to predict ranges of expected performance and yield for given specifications. Statistical modeling can then be used as a guide for redesign, in order to improve yield. Statistical modeling can also be used to study the mismatch between identically laid out devices; such studies are very important for analog circuit design.¹⁰²⁻¹⁰⁶

It should be stressed that some of the advantages of physical models just mentioned might not be valid if the model parameters have not been chosen correctly. This issue will be discussed later in this chapter.

Good models are usually complex since they have to take into account carefully several phenomena which the simpler models ignore. By the time all regions of operation are included, one can easily end up with models with, say, 40 parameters. These can provide accuracy at the expense of complexity. Note however that satisfactory accuracy and a large number of parameters do not always mean that the physics of the device has been modeled correctly. There exist models which are based on faulty premises but which, just because they contain many parameters, can give adequate accuracy after extensive empirical adjustment of their values. Such models, though, do not normally have predictive power.

10.3 COMBINING SEVERAL EFFECTS INTO ONE PHYSICAL MODEL

In previous chapters, we have discussed many different phenomena that take place in the MOS transistor, and have considered ways to model them. For the purposes of presentation, our approach was to focus on one phenomenon at a time. A complete device model, though, must incorporate all these phenomena simultaneously. Doing so properly would require considering the interaction between these effects, but then, unfortunately, the expressions obtained become very complicated. An empirical approach often taken, *if each effect by itself is small*, is to assume that the effects are noninteracting. For example, consider a device with a channel that is both short *and* narrow. Here three-dimensional analysis would be needed to obtain accurate results. However, for simple estimates such a device is sometimes modeled by using an effective-threshold formulation, with the effective threshold \hat{V}_T being given by $V_T + \Delta V_{TL} + \Delta V_{TW}$, where V_T is the long-channel threshold, ΔV_{TL} is the change assuming only short-channel effects are present, and ΔV_{TW} is the change assuming only narrow-channel effects are present. For small ΔV_{TL} and ΔV_{TW} , this approach can be justified to some extent (Prob. 6.17).

In developing certain models, one is faced not only with combining several effects that may be active in a given region of operation, but also with combining different expressions that have each been developed for particular regions only (e.g., combining nonsaturation and saturation expressions). In such cases, one must make sure that, as a boundary between two regions is crossed, continuity is maintained not only for I_{DS} but also for its derivatives with respect to the terminal voltages.

We now illustrate the above considerations with a simple example.

Example 10.1. Let us attempt to develop drain current expressions which include the following effects:

1. Effect of L on effective threshold
2. Effect of V_{DS} on effective threshold
3. Effect of W on effective threshold
4. Velocity saturation effects
5. Effective mobility dependence on normal field
6. Channel length modulation in saturation

The approach used will be a variation of one used elsewhere.²¹ Effects 1, 2, and 3 will be modeled by using the following effective threshold voltage:

$$\hat{V}_T(L, W, V_{DS}, V_{SB}) = V_T(V_{SB}) + \Delta V_{TL}(L, V_{DS}, V_{SB}) + \Delta V_{TW}(W, V_{SB}) \quad (10.3.1)$$

where $V_T(V_{SB})$ can be calculated from (4.5.32), and ΔV_{TL} and ΔV_{TW} can be calculated as described in Sec. 6.3.

Effect 4 will be modeled as in Example 6.2 (Sec. 6.5), and effect 5 will be included by using (4.10.20) [see discussion following (6.5.11)]. Thus, in the nonsaturation region we have

$$I_{DS} = \frac{\mu_0 C'_{ox} \frac{W}{L} \left\{ [V_{GS} - \hat{V}_T(L, W, V_{DS}, V_{SB})] V_{DS} - 0.5 \alpha V_{DS}^2 \right\}}{\left\{ 1 + \theta [V_{GS} - \hat{V}_T(L, W, V_{DS}, V_{SB})] + \theta_B V_{SB} \right\} \left[1 + V_{DS}/(L\mathcal{E}_c) \right]}, \quad V_{DS} \leq V \quad (10.3.2)$$

where \hat{V}_T is given by (10.3.1).

The value of the saturation drain-source voltage V'_{DS} will be found as shown in Sec. 6.5 by setting dI_{DS}/dV_{DS} equal to zero from (10.3.2) and neglecting the dependence of V_T on V_{DS} for this calculation. This results in (6.5.12) again. A more exact calculation results in a much more complicated expression which gives practically the same value as (6.5.12).²¹

We will now determine I_{DS} in saturation. This requires some caution. If \hat{V}_T were independent of V_{DS} , as implicitly assumed in Sec. 6.2, we could use (6.2.5), where I'_{DS} would be given from (10.3.2) after replacing V_{DS} by V'_{DS} . In the resulting expression for I'_{DS} , V_{DS} itself would not appear. This, of course, would be so because the channel end, being considered to be pinched off, would be assumed to be at potential V'_{DS} with respect to the source, no matter what the actual value of V_{DS} . However, here we want to include the effect of V_{DS} on \hat{V}_T , as explicitly indicated in (10.3.1). As already pointed out, this effect is assumed unrelated to pinchoff and is present whether V_{DS} is smaller or larger than V'_{DS} , as explained in Sec. 6.3. Hence, even in saturation, \hat{V}_T will continue to be a function V_{DS} , not V'_{DS} . Accordingly, we have

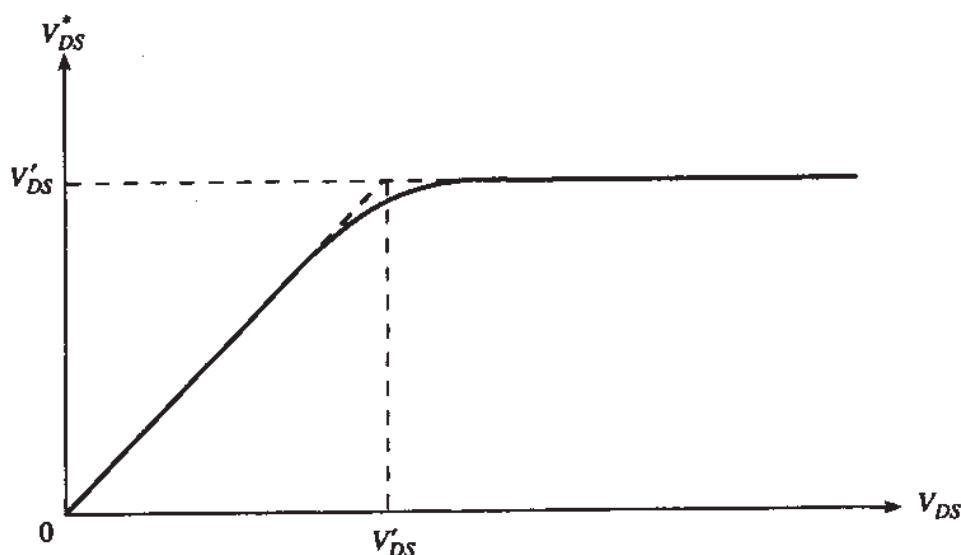
$$I_{DS} = \frac{\mu_0 C'_{ox} \frac{W}{L} \left\{ [V_{GS} - \hat{V}_T(L, W, V_{DS}, V_{SB})] V'_{DS} - 0.5 \alpha V_{DS}^2 \right\}}{\left\{ 1 + \theta [V_{GS} - \hat{V}_T(L, W, V_{DS}, V_{SB})] + \theta_B V_{SB} \right\} \left[1 - \frac{l_p}{L} + V'_{DS}/(L\mathcal{E}_c) \right]}, \quad V_{DS} > V'_{DS} \quad (10.3.3)$$

with l_p as in Sec. 6.2. Note that, with the exception of the argument of \hat{V}_T , V_{DS} has been replaced by V'_{DS} in the above equation.

We now need to make sure that the nonsaturation and saturation expressions give not only the same value of I_{DS} at $V_{DS} = V'_{DS}$, but also the same slope dI_{DS}/dV_{DS} . This can be achieved by a slight modification of the value of V'_{DS} , as explained in Sec. 6.2. However, there are still two problems. First, although the redefinition of V'_{DS} can guarantee continuity in dI_{DS}/dV_{DS} , it does not guarantee continuity in $d^2I_{DS}/dV_{DS}^2 = dg_{sd}/dV_{DS}$. This results in a nonphysical behavior of g_{sd} . Second, the use of (10.3.2) or (10.3.3) depending on whether or not $V_{DS} \leq V'_{DS}$, requires an "IF" statement in the computer code implementing the model. This is undesirable (Sec. 10.6).

The above problems would not exist if a more detailed modeling had been done, using the charge sheet models of Sec. 4.3. These make possible a single expression for I_{DS} , which saturates smoothly, with g_{sd} also varying smoothly, as V_{DS} is increased. If such a model cannot be used, an approach sometimes taken is the use of "smoothing functions."⁴⁴ According to this approach, we can replace (10.3.2) and (10.3.3) by the following single-piece expression:

$$I_{DS} = \frac{\mu_0 C'_{ox} \frac{W}{L} \left\{ [V_{GS} - \hat{V}_T(L, W, V_{DS}, V_{SB})] V_{DS}^* - 0.5 \alpha V_{DS}^{*2} \right\}}{\left\{ 1 + \theta [V_{GS} - \hat{V}_T(L, W, V_{DS}, V_{SB})] + \theta_B V_{SB} \right\} \left[1 - \frac{l_p}{L} + V_{DS}^*/(L\mathcal{E}_c) \right]} \quad (10.3.4)$$

**FIGURE 10.1**

"Effective" drain-source voltage V_{DS}^* as a function of actual drain-source voltage V_{DS} .

where V_{DS}^* is a quantity that approaches V_{DS} if $V_{DS} < V_{DS}'$, and V_{DS}' if $V_{DS} > V_{DS}'$, as shown in Fig. 10.1. Of course, the transition from $V_{DS}^* = V_{DS}$ to $V_{DS}^* = V_{DS}'$ should be appropriate for providing realistic values for the current and its slope. An empirical function used for this purpose is^{37,44}

$$V_{DS}^* = V_{DS}' \left\{ 1 - \frac{\ln \left[1 + e^{P(1 - V_{DS}/V_{DS}')} \right]}{\ln \left[1 + e^P \right]} \right\} \quad (10.3.5)$$

where P can be chosen according to the sharpness desired for the transition region (e.g.,³⁷ $P = 10$).

We note that we have already seen the use of such smoothing or interpolation functions for joining the weak and strong inversion regions.^{28,61}

The above is meant only as a simple example for illustrating some common approaches in modeling. Note that the model is clearly incomplete, as is discussed in earlier chapters; notably, it completely ignores the moderate- and weak-inversion regions. It is also emphasized that empirical smoothing functions should only be used as a last resort. As MOSFET modeling advances, it is hoped that popular models will be increasingly based on general expressions such as those of the charge sheet model in Sec. 4.3; such expressions, by correctly incorporating the physics of the transistor, exhibit continuity for I_{DS} and its derivatives for any combination of terminal voltages. Also, more accuracy would be possible above if the interaction between the various effects were considered, but this would complicate the equations. Further complexity is needed in order to take into account more effects, or to model the above effects in more detail. Indeed, to describe and attempt to justify even one of the more complete compact models would take a large part of this book. This is why such models will not be presented in detail here. The interested reader is referred instead to the literature; see, for example, Ref. 65.

10.4 PARAMETER EXTRACTION

No matter how sound a physical model is, it cannot give accurate results unless appropriate values are used for its parameters. These values should be chosen so that the model predicts a behavior as close as possible to measurements, as illustrated for I_{DS} in Fig. 10.2. Determining these values^{107-206,38,47,50} is not a simple matter, for several reasons. First, the value of some of these parameters may not be known accurately; for example, the flat-band voltage may not have been measured independently. Second, some of the parameters in the model are, as we have seen, basically empirical in nature, which means by definition that they should be chosen for best matching to measurement. Third, even if the value of a physical parameter is known accurately, this value may not be the best one to use in the model expressions. This is because analytical physical models are based on several assumptions and approximations; we have seen many instances of this fact in previous chapters. Thus, using these expressions with the “correct” values for their parameters results in a certain error. By slightly modifying the value of these parameters, the error can often be decreased; we have already discussed this fact in Sec. 4.15. In other words, a slight modification of the parameter values can in some instances partially make up for model deficiencies. Unfortunately, this is often abused, as we will discuss later on in this section.

A common way to determine a parameter value (or, at least an initial guess for it) is to concentrate on a particular region of operation where this parameter has a dominant effect, and obtain enough data so that the parameter value can be deduced. Often, it is attempted to identify regions where a first-degree polynomial relation is

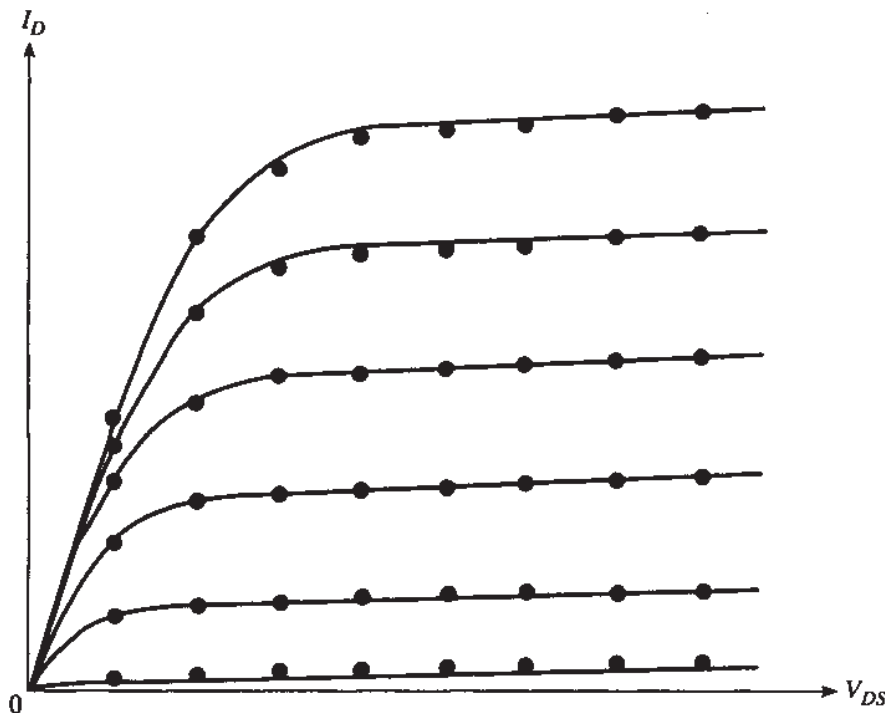


FIGURE 10.2

I_D versus V_{DS} , with V_{GS} a parameter. Points: measurements; lines: model.

supposed to exist between an independent variable (e.g., an externally applied voltage) and a dependent one (e.g., a measured current); the parameter values are chosen so that the model matches that relation with minimum error. This is done using an optimization technique known as *linear regression*,²⁰⁷ which is standard in several mathematics and data handling packages; for the purposes of visualization, we will show this process as fitting a straight line through data on I - V plots.

As an example, a common method to determine several parameters is to use data in strong inversion, with V_{GS} of at least 0.5 V above V_T , with very low V_{DS} (e.g., 50 mV). Assume initially that the extrinsic resistances of the source and drain regions are zero, for simplicity. Since V_{DS} is very small, the drain-body current can be neglected and thus $I_D = I_{DS}$. From (4.5.37a) we have

$$I_D = \frac{W}{L} \mu C'_{ox} \left(V_{GS} - V_T - \frac{\alpha}{2} V_{DS} \right) V_{DS} \quad (10.4.1a)$$

$$\approx \frac{W}{L} \mu C'_{ox} (V_{GS} - V_T) V_{DS}, \quad V_{DS} \ll V_{GS} - V_T \quad (10.4.1b)$$

Measured I_D values corresponding to this situation are shown versus V_{GS} in Fig. 10.3, for a given V_{SB} . As seen, the behavior deviates from the one expected from (10.4.1); this is because at very low V_{GS} the device operates in moderate inversion, where (10.4.1) is not valid; and, at high V_{GS} , the effective mobility decreases as V_{GS} is increased (Sec. 4.10). Nevertheless, one hopes that there will be a region of V_{GS} where (10.4.1) will be approximately valid with a constant μ ; this will be the region of highest slope in the figure. Extrapolating from that region as shown¹⁹² down to

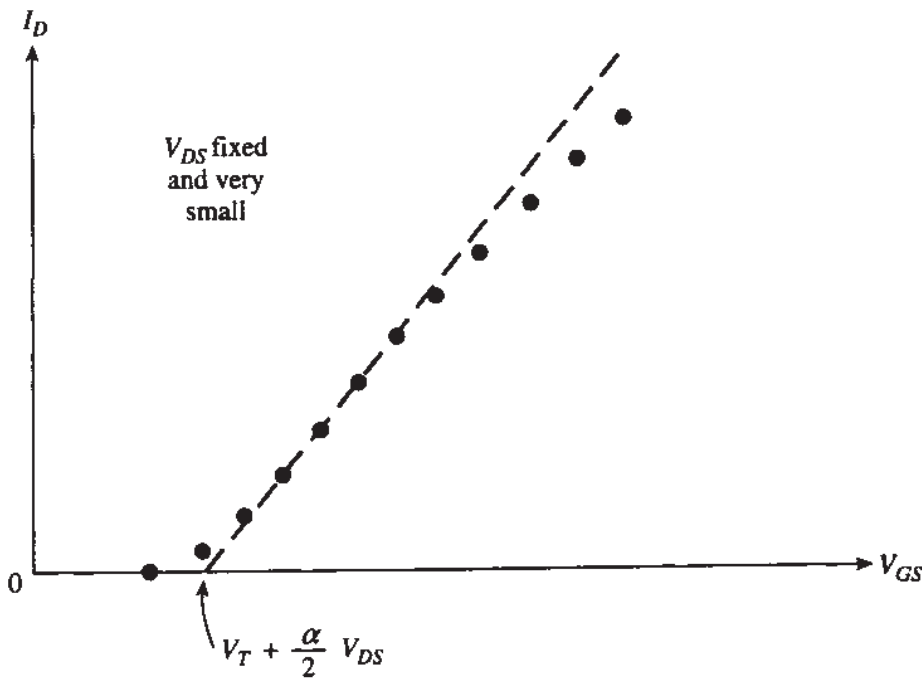


FIGURE 10.3

I_D versus V_{GS} , for a fixed, very small V_{DS} (e.g., 50 mV) and fixed V_{SB} . Points: measurements. The broken line is used to extrapolate from the region of maximum slope.

zero I_D gives an intercept of approximately V_T , as expected from (10.4.1b); more precisely, the intercept is $V_T + (\alpha/2)V_{DS}$, as seen from (10.4.1a). Using then an estimate for α , V_T can be found from the above intercept (the error due to the fact that α is not known accurately will be small, since V_{DS} is chosen very small for this measurement).[†]

The above process can be repeated for several values of V_{SB} , thus determining the behavior of $V_T(V_{SB})$. This can then be used to determine the parameter values for a given model of this behavior. For example, if it is assumed that the latter follows the simple model of (4.5.32), we have $V_T = V_{FB} + \phi_0 + \gamma\sqrt{V_{SB} + \phi_0}$; thus, V_T should be a straight line with respect to $\sqrt{V_{SB} + \phi_0}$. The parameters ϕ_0 , V_{FB} , and γ can be chosen to fit this behavior (Prob. 10.2). Capacitance measurements are also used to determine V_{FB} .

The quantities W and L used in (10.4.1) and in all equations for I_{DS} in this book are the “electrical” channel width and length, which are different from the corresponding “mask” (or “drawn”) values, W_m and L_m respectively, specified during the layout process. The electrical and mask values are related by (1.6.1) and (1.6.2), repeated here:

$$W = W_m - \Delta W \quad (10.4.2)$$

$$L = L_m - \Delta L \quad (10.4.3)$$

The values of W_m and L_m are known from layout information; however, ΔW and ΔL depend on several variables during fabrication and are not known accurately. To determine ΔW from measurements,³⁸ note that, assuming that V_{DS} is very small, the conductance $G = I_D/V_{DS}$ is, from (10.4.1b) and (10.4.2),

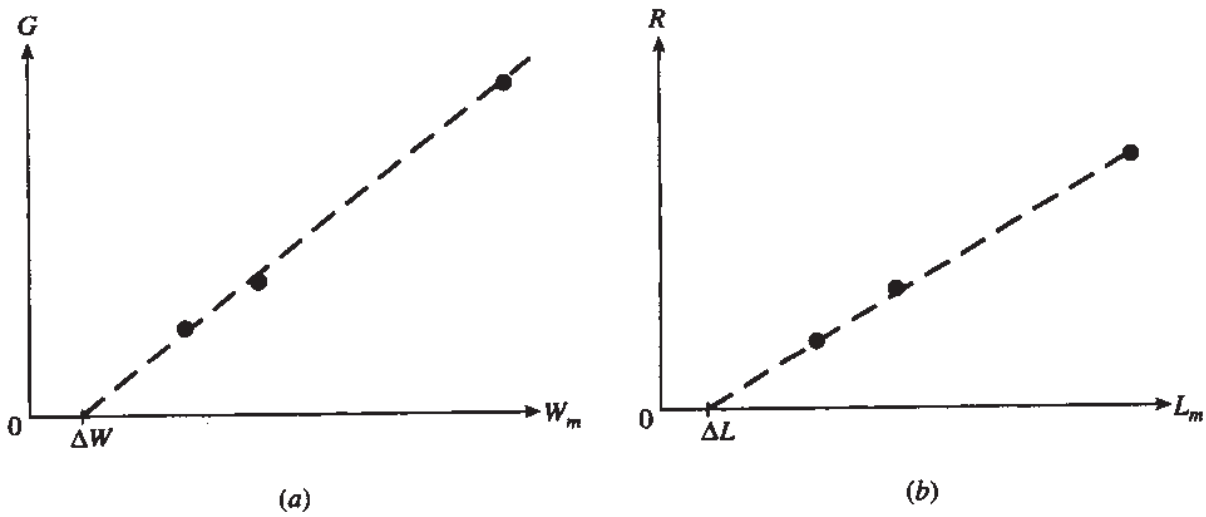
$$G = \frac{I_D}{V_{DS}} = \left[\frac{\mu C'_{ox}(V_{GS} - V_T)}{L} \right] (W_m - \Delta W) \quad (10.4.4)$$

This quantity is thus expected to be a straight line versus W_m . If G is measured for several devices, each with the same L_m but with different W_m values, we obtain the situation shown in Fig. 10.4a. Using linear regression, a best-fit first-degree polynomial can be determined. Extrapolating to zero G gives an intercept equal to ΔW , according to (10.4.4). To avoid interference from short-channel effects in this procedure, the common value of L_m used for the measurements should be chosen large.

A similar procedure³⁸ can be followed to determine ΔL . Again assuming a very small V_{DS} , the resistance $R = V_{DS}/I_D$ is, from (10.4.1b) and (10.4.3):

$$R = \frac{V_{DS}}{I_D} = \left[\frac{1}{W\mu C'_{ox}(V_{GS} - V_T)} \right] (L_m - \Delta L) \quad (10.4.5)$$

[†]Another common technique¹⁹¹ estimates V_T as the value of V_{GS} needed to produce a certain current (commonly $0.1\mu A \times (W/L)$ in weak inversion. This is referred to as *constant current threshold*. Yet another technique uses the V_{GS} intercept of a plot of $\sqrt{I_D}$ versus V_{GS} in saturation [see (4.5.37b)]; however, DIBL and velocity saturation can make the results inaccurate.

**FIGURE 10.4**

(a) Strong-inversion channel conductance versus mask width (points), and extrapolation (broken line) for determining ΔW . (b) Channel resistance versus mask length (points), and extrapolation (broken line) for determining ΔL . The source and drain extrinsic resistance are assumed negligible. V_{GS} , V_{SB} , and V_{DS} are fixed; V_{DS} is very small.

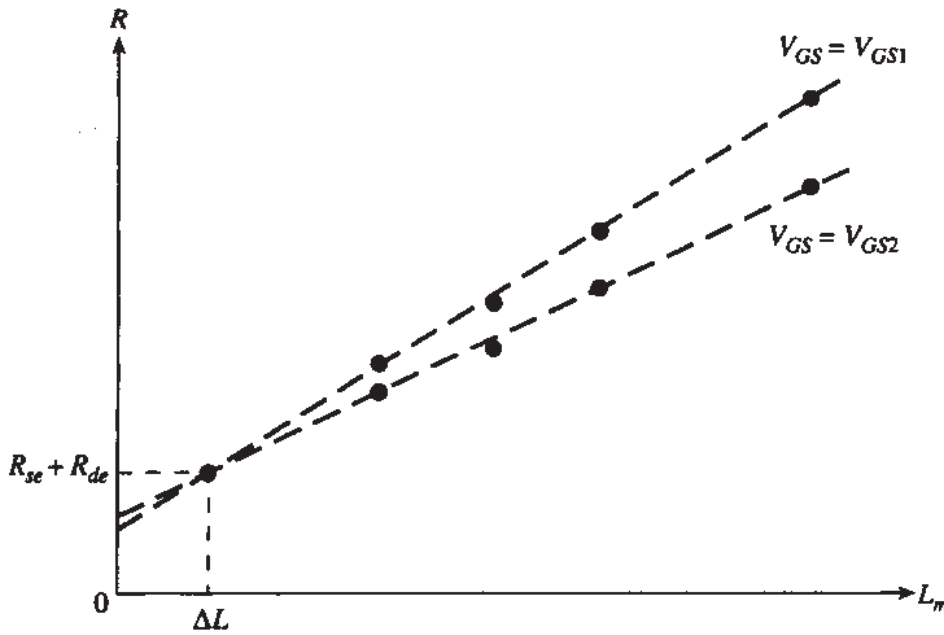
and is thus expected to be a straight line when plotted versus L_m . If one determines R for several devices, all with the same large W_m , but with different L_m , one obtains the behavior shown in Fig. 10.4b. Using linear regression again, and extrapolating to zero R , gives an intercept equal to ΔL .

It is clear that the above procedures will give only rough estimates for the parameter values, since several effects are neglected. For more accurate parameter determination, these techniques have to be refined. Consider, for example, the effect of source and drain extrinsic resistances, R_{se} and R_{de} . If these are included, the total resistance seen between source and drain will be given by, instead of (10.4.5),

$$R_{\text{tot}} = R_{se} + R_{de} + \left[\frac{1}{W\mu C'_{\text{ox}}(V_{GS} - V_T)} \right] (L_m - \Delta L) \quad (10.4.6)$$

To determine both ΔL and $(R_{se} + R_{de})$, one can consider again devices of different L_m values, and plot the measured R_{tot} versus L_m . If this is done for at least two V_{GS} values, a linear regression is done for each V_{GS} value, and the results are extrapolated as shown in Fig. 10.5, the resulting lines will cross at $L_m = \Delta L$, and the value of the ordinate at this point will give $R_{se} + R_{de}$, as seen from (10.4.6).¹⁸⁹ Thus, this procedure gives both ΔL in the presence of a nonzero series resistance, and the value of that resistance itself. A similar refinement is possible for determining ΔW in the presence of nonzero series resistance (Prob. 10.4).

Further refinements may still be needed. For example, in LDD devices the lightly doped part of the source and drain regions gives series resistances which are functions of bias. Thus the two V_{GS} values used in Fig. 10.5 should be close to each other,^{138,140,154} and the value of $R_{se} + R_{de}$ determined in this way would be acceptable for $V_{GS} = (V_{GS1} + V_{GS2})/2$. Different V_{GS} values would result in different $R_{se} + R_{de}$. In

**FIGURE 10.5**

Strong-inversion channel resistance versus mask length (points) for two values of V_{GS} , and extrapolations (broken lines) for determining ΔL and $R_{se} + R_{de}$. V_{GS} , V_{SB} , and V_{DS} are fixed; V_{DS} is very small.

fact, even ΔL turns out to be a function of V_{GS} for LDD devices. In this case, ΔL can be interpreted to be an “effective ΔL ,” such that simple model equations (derived assuming heavily doped source and drain regions) give results which are approximately valid for LDD devices.

With W and L determined using the above procedures, one can return to (10.4.1b) in order to determine the effective mobility. The low-field mobility μ_0 (i.e., when V_{GS} is not large enough to reduce the mobility value) is sometimes estimated as that corresponding to the maximum slope in Fig. 10.3. In other words, it is assumed that in the region of maximum slope μ is equal to μ_0 , and thus the slope of the plot is $g_m = \partial I_D / \partial V_{GS} \approx (W/L)\mu_0 C'_{ox} V_{DS}$ [see (10.4.1b)]. Thus, the measured g_m is plotted versus V_{GS} , its maximum $g_{m,max}$ is found, and μ_0 is estimated by $g_{m,max} / [(W/L)C'_{ox} V_{DS}]$, where W and L are found from (10.4.2) and (10.4.3), and C'_{ox} is assumed known from capacitance measurements. Clearly, this process can fail in cases where the mobility dependence on V_{GS} is strong, as then there may not be any region where $\mu = \mu_0$ in the plot of Fig. 10.3. In that case, $\mu = \mu(V_{GS})$ in (10.4.1b), and thus one cannot assume μ is a constant when differentiating with respect to V_{GS} , as was done above. It is better⁴⁷ to differentiate I_D in (10.4.1b) with respect to V_{DS} , which gives $g_o = \partial I_D / \partial V_{DS} = (W/L)\mu(V_{GS})C'_{ox}(V_{GS} - V_T)$. Thus:

$$\mu(V_{GS}) = \frac{g_o}{(W/L)C'_{ox}(V_{GS} - V_T)} \quad (10.4.7)$$

This can be used provided V_{GS} is restricted to strong inversion, where the relations used above are valid. Using measurements of g_o at several values of V_{GS} , one can thus determine the behavior of μ versus V_{GS} . From this, one can find the constants used in the effective mobility expression (Sec. 4.10 and Prob. 10.10).

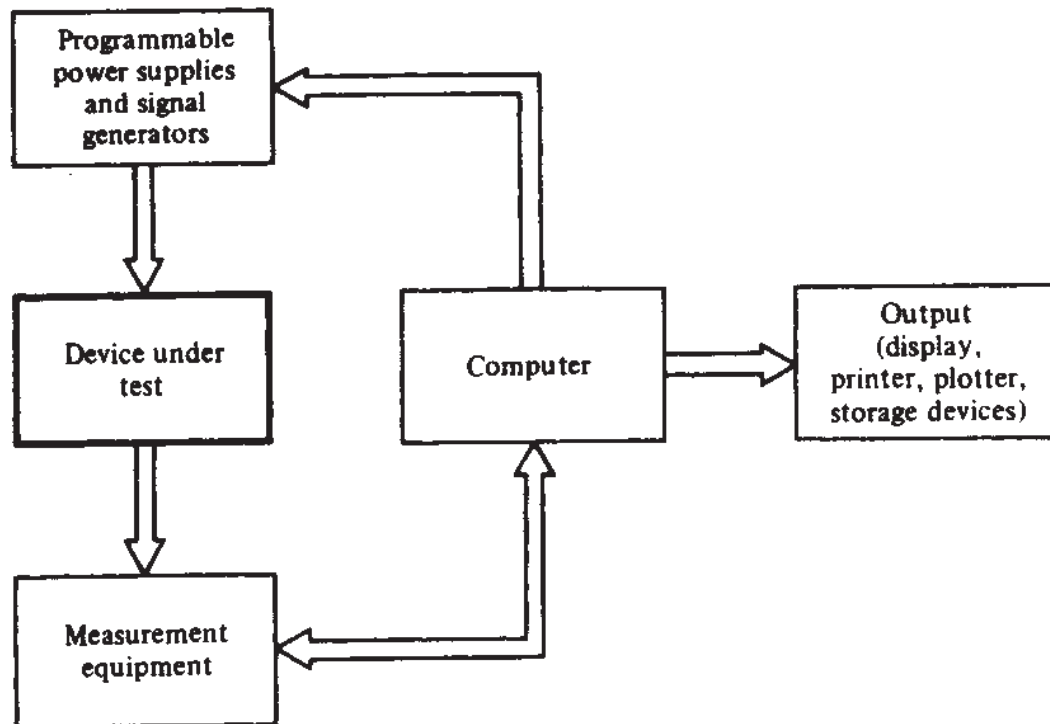


FIGURE 10.6
A system for data acquisition and parameter extraction.

The techniques mentioned above are only meant to be illustrative examples; they are not the only techniques used. A variety of approaches are employed with various degrees of sophistication, depending on the model used, the accuracy desired, etc.^{107-206,47,50}

The process of determining values for the parameters in a model is called *parameter extraction* or *characterization* and is combined with measurement, as shown in Fig. 10.6. A large number of different bias voltages and signal excitations are applied to the transistor under test, by programmable test sources under the control of a computer. The response of the device (e.g., the corresponding drain current values) is measured by programmable equipment, also under computer control, and stored in memory. An *optimizer* program¹⁰⁷⁻¹³³ then determines the best parameter values to use for a given model, so that a certain criterion is met; for example, it is usually desirable to choose the parameter values so that the I_{DS} values predicted by the model match the measured values with as little overall error as possible, where “error” is defined in some appropriate way to be discussed below. This is illustrated in Fig. 10.2.

In some cases, the data to be matched may come not from a measurement system, but from a device simulator. This would in fact be necessary if it is desired to produce an analytical model before actual fabricated devices are available for measurements.

ERROR CRITERION. The error to be minimized during the optimization procedure is often defined as follows:

$$E_I = \sum_{j=1}^K w_j \left(\frac{\hat{I}_{Dj} - I_{Dj}}{I_{Dj}} \right)^2 \quad (10.4.8)$$

where I_{Dj} , $j = 1, \dots, K$, are measured I_D values for K different combinations of bias voltages, \hat{I}_{Dj} are the corresponding calculated values, and w_j are weighting coefficients, which are often chosen equal to 1; if desired, these coefficients can be chosen larger in certain regions of operation, in which case the error in those regions will count more in the overall error, and the parameter extraction routine will thus be forced to “pay more attention” to those regions. The squares are used so that both positive and negative errors count equally, and so that such errors do not cancel each other out; sometimes, absolute values are used instead of squares. Note that, for very small currents, the denominators in (10.4.8) become very small, whereas, in general, the errors represented by the numerators do not. This tends to assign more importance to the error at small current values, and can be corrected by choosing the weighting coefficients accordingly. If no weighting coefficients are used, it may be desirable not to allow the denominator to decrease below a certain value, say I_{\min} ; it is thus common to replace the denominators in (10.4.8) by $\max(I_{Dj}, I_{\min})$.

The root mean square (rms) error corresponding to (10.4.8) is the square root of its mean value:

$$\epsilon = \sqrt{\frac{E_I}{K}} \quad (10.4.9)$$

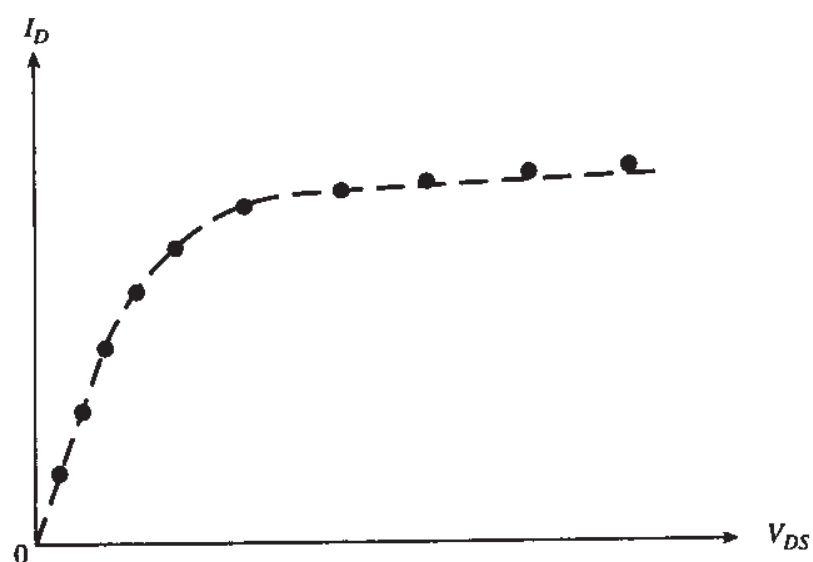
For analog applications, minimizing just (10.4.8) does not necessarily produce adequate results. Let us revisit a point we made in Chap. 8. Assume that a parameter extraction for a certain model has been done in order to minimize the above error, resulting in the satisfactory fit to data shown in Fig. 10.7a. When the same model is used to predict the small-signal output conductance g_o (the derivative $\partial I_D / \partial V_{DS}$), the result is as in Fig. 10.7b. The error involved in predicting this parameter can be defined in a similar way as above:

$$E_g = \sum_{j=1}^K u_j \left(\frac{\hat{g}_{oj} - g_{oj}}{g_{oj}} \right)^2 \quad (10.4.10)$$

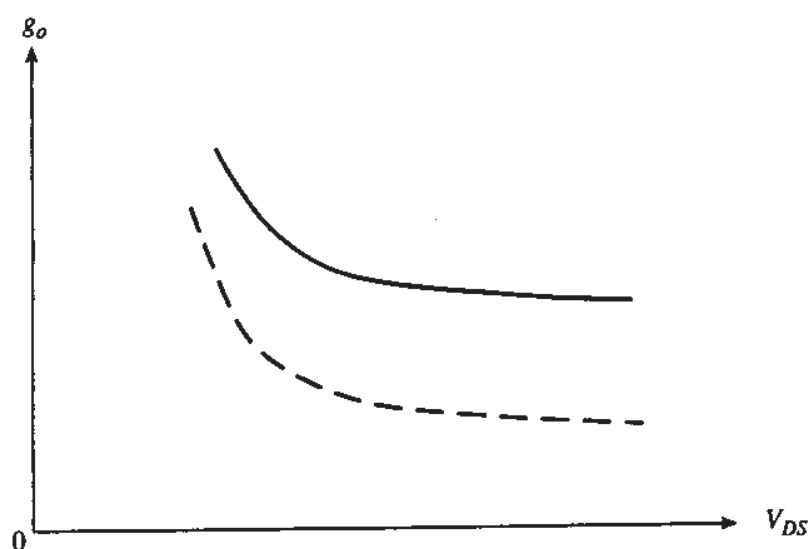
where u_j are weighting coefficients. As seen in Fig. 10.7b, this error can be very large. This is because the slope in saturation is so small that, even if it is in error by, say, a factor of 2, the corresponding effect in the I_D curve (Fig. 10.7a) will be very small and will not be “noticeable,” if minimizing E_I is the objective. To rectify this situation, the optimizer should be asked instead to optimize a combination of E_I and E_g above;²⁰⁸ a simple combination is the sum of the two errors:

$$E = w_0 E_I + u_0 E_g \quad (10.4.11)$$

where w_0 and u_0 are weighting coefficients, although other ways to perform optimization with multiple objectives do exist. With some models, use of the above criterion can provide a drastic improvement in g_o accuracy, with negligible loss in overall current accuracy.^{208,209,120,27} One must, of course, set up the appropriate facilities for measuring g_o ; this measurement can be difficult.²⁰⁶ Errors involving g_m and g_{mb} can also be used. However, these parameters can be predicted relatively accurately if the current is predicted accurately.



(a)



(b)

FIGURE 10.7

(a) I_D - V_{DS} characteristic. Points: measurements; broken line: model. (b) Slope g_o corresponding to (a). Solid line: measurement; broken line: model.

The above errors involve only dc measurements. Ideally, one could envision parameter extraction in which the error in other quantities (ac response, rise times, noise, etc.) could be taken into account. However, this involves difficult measurements and greatly complicates parameter extraction. Thus, in most cases the minimization involves only dc quantities, and often even the errors in g_o are ignored, to the distress of analog circuit designers.

EXTRACTION STRATEGY. It is in principle possible to give to an optimizer program a general model involving all parameters (say, 40 of them) and ask it to extract the

parameter values to minimize an appropriate error as above, without further guidance. This is, however, numerically very complicated. Also, even if the process converges and a small error results, the parameter values it ends up with may not make physical sense. The optimizer does not understand physics; it treats the entire process as a curve-fitting exercise. Any parameter value that helps the fitting is “good,” as far as the optimizer is concerned. One may well ask, “What is the problem with this approach? Isn’t good fitting what we are after, anyway?” The answer to this question is that we are not only after fitting of the data fed to the optimizer. We hope that, once the optimizer finds parameter values with the help of the data provided, these values will be good for also making predictions of the current in a bias range not adequately represented in the measured data, or of other quantities altogether, for example the device capacitances. There is no guarantee that such hopes will actually be fulfilled with the above approach.

Let us provide an absurd example, to show the dangers of “blind” optimization. Consider the quantity ϕ_0 in the expression for the threshold voltage, (4.5.32). This quantity is often assumed to be given by $\phi_0 = 2\phi_F = 2\phi_i \ln(N_A/n_i)$ [see (2.5.22) and (1.4.2)]. However, close examination reveals that the effective value of ϕ_0 can differ from $2\phi_F$ by several ϕ_F . This is true for implanted devices [see (5.2.22)], and even for uniform substrates [see (2.5.23)]. If, during parameter extraction, we make the mistake to view N_A as a parameter to be “extracted” from the above formula, the optimizer will try different values for N_A until ϕ_0 attains the value required to predict the body effect, and thus the variation of I_{DS} with V_{SB} , adequately. Because of the logarithmic form of the above relation, to make up for a small error involved in defining ϕ_0 , N_A will have to be given a very artificial value (about 7 times higher, for a ϕ_0 error of 100 mV). This artificial value may interfere with the proper prediction of other phenomena in which N_A plays a role, and which may not have been adequately represented in the characterization procedure.

The above example makes obvious the fact that blind parameter extraction is a very dangerous process. Thus, the optimizer is usually guided to produce physically sensible results. A judicious choice is made as to which parameters are to be extracted, as opposed to those which can be set from process knowledge (the latter can include, for example, the oxide thickness). Then, the set of parameters to be extracted is separated into several subsets, each of which can be extracted independently of the others, by concentrating on that part of the device characteristics which is most strongly determined by that group of parameters. For example, measured data in the nonsaturation region with low V_{DS} (e.g., 50 mV) is used to extract the values of parameters strongly affecting this region (e.g., those related to the effective channel length and width, effective mobility, threshold voltage, and the body effect). With the values of these parameters set, data in the strong inversion saturation region is used to extract the values of parameters affecting channel length modulation, velocity saturation, etc. Then weak inversion data is used to extract relevant weak inversion parameters, etc. The measured data is chosen appropriately in each region corresponding to each parameter subset, to facilitate proper extraction in that region. A good initial guess is provided for the values of the various parameters to be extracted. The optimization performed is suitably constrained, with the allowable range of values for

each parameter specified, thus helping the optimizer converge to physically meaningful values. After all parameter values have been determined, sometimes the optimizer is allowed to do a fine tuning on all of them, considering the entire set of measured data. This may help, since some parameters affect more than one region, and this global optimization can take this into account.

The particular way to extract each parameter depends on the parameter itself and on the way this parameter appears in the model. Many ways to measure and extract various parameter values have been discussed in the literature.^{107-206,47,50} Several test devices, of various gate lengths and widths, are included on a wafer for measurement and extraction purposes.

A good model will have enough parameters so that all important phenomena can be modeled, but on the other hand it will not include redundant parameters, i.e., parameters to which the device characteristics are not sensitive, or parameters which can cancel each other's effect. And, most importantly, *if the model is good, the extracted parameter values will be close to what is expected from physics.*

BINNING. Ideally, to obtain results from a model one should only have to provide the channel width and length, the temperature, and one and the same set of parameter values extracted as discussed above. In reality, most models are not that good; if their parameters are extracted by fitting to measurements on a device with, say, $W = 1\ \mu\text{m}$ and $L = 0.35\ \mu\text{m}$, these parameters will not be adequate for modeling a device made in the same process with significantly different dimensions. This means that the model equations do not do an adequate job in taking the effect of device geometry into account. In such cases, to make up for the model deficiencies, a process known as "binning" is used.⁶⁵ The expected range of W and L values is divided up into "bins" (e.g., a bin may be defined by $0.35\ \mu\text{m} \leq L < 0.5\ \mu\text{m}$ and $0.7\ \mu\text{m} \leq W < 1.5\ \mu\text{m}$), and parameter values are extracted for a representative device in each bin. This necessitates the use of different sets of parameters during circuit design, depending on the desired L and W . The danger with this scheme is that, if a dimension happens to be at or near the edge of a bin, a slight change of this dimension by the designer might produce a large change in simulated circuit performance. In addition, during statistical simulations, in which the random variation of L and W is considered, crossing the boundary between bins can produce complications.

10.5 ACCURACY

It is often said that all that need be expected of a model (with its associated parameter extraction) is to predict the current throughout the bias ranges of interest with an accuracy of 5 to 10 percent, since the statistical variations in the process parameters imply an uncertainty much larger than 10 percent anyway. Whether this is a valid claim or not depends on the application. For most of digital circuit design, it may be true. In certain analog circuits, though, what counts is not only the values of the current, but also the form of its functional dependence on the bias voltages in certain ranges and the corresponding small-signal parameters. The model used in Fig. 10.7a

is accurate within 5 percent in terms of the current, yet it totally fails to predict g_o decently, as seen in Fig. 10.7b; this can have serious consequences in circuit design, where amplifier gains are often inversely proportional to sums of g_o quantities. In cases where g_o is of prime importance, slope accuracy is more important than current accuracy! Also, some models produce artificial kinks in the I - V characteristics, which can seriously interfere with analog circuit design; yet such models can be accurate to 5 to 10 percent.

In conclusion, at least for simulation of analog and mixed analog/digital circuits, a current accuracy within the expected statistical spread is not enough. We also need a *small-signal parameter accuracy* within the corresponding expected statistical spread, and a correct functional dependence and qualitative behavior. A good physical model can be counted on to satisfy these requirements, if its parameters have been determined well.

10.6 PROPERTIES OF GOOD MODELS

Here is a summary of requirements which a CAD model must ideally meet to be suitable for analog and mixed analog/digital circuit design:⁵⁴

1. The model should, of course, meet common requirements for digital work, such as reasonable accuracy in I - V characteristics and extrinsic parasitics.
2. It should give accurate values for all small-signal quantities such as g_m , g_{mb} , g_{sd} , and capacitances. In particular, all these parameters should be continuous with respect to any terminal voltage.
3. It should give good results even when the device operates non-quasi-statically, or at least it should degrade gracefully for such operation, as frequency is increased.
4. It should give accurate predictions for both white and $1/f$ noise, including in the triode region.
5. It should meet requirements 1 to 4 above over large bias ranges, including $V_{SB} \neq 0$, and encompassing the weak-, moderate-, and strong-inversion regions.
6. It should do all of the above over the temperature range of interest.
7. It should do all of the above for any combination of channel width and length values, from the minimum specified upward.
8. The user should only have to specify the geometrical dimensions for each device, the temperature, and one set of model parameters valid for all devices of the same type independent of dimensions.
9. The model should provide a flag every time it is attempted to use it outside its limits of validity. For example, if the model is quasi-static and one attempts to use it, say, above the unity-gain frequency of the device, a warning should be given to the user that the result may be inaccurate.
10. The model should have as few parameters as possible (but just enough), and those parameters should be linked as strongly as possible to ones related to the

device structure and fabrication process (e.g., oxide thickness, substrate doping, ion implantation characteristics, junction depth). This would allow meaningful statistical and worst-case simulations and yield predictions, as well as studies of mismatches between devices (which are important in analog design). An emphasis on process and geometrical parameters is also necessary in order to make the model predictive and to make it easily adaptable in the face of process and layout rule changes. Empirical parameters without physical meaning should be avoided as much as possible.

11. The model should be computationally efficient and numerically robust.
12. It should allow for the modeling of asymmetric devices; for example, it should allow the freedom of specifying junction capacitance parameters separately for the source and drain regions.⁷⁰ This is needed not only because a variety of layouts are used by designers, but also because in some technologies (notably high-voltage ones) different fabrication steps are used for the source and drain regions.
13. The model should be linked to an efficient parameter extraction method. The number of required test devices and tests for parameter extraction should be as small as possible.
14. The model should ideally provide links to device simulators.

If the performance of present models is judged against the above criteria, it becomes obvious that MOSFET modeling still has a long way to go.

10.7 CONSIDERATIONS AND CHOICES

10.7.1 General Considerations

In order to make the above wish list come true, several considerations and choices confront a model developer. Some of these are now discussed.

1. To begin with, the only type of model that can meet the above set of requirements is a physical model. The advantages of physical models have already been discussed above.
2. A lot of thought must be given to the choice of primary parameters for the model. For example, V_{T0} should not be such a parameter, since this would hide the dependence of threshold on oxide thickness, which may be important in statistical simulation;⁷⁰ if, instead ϕ_{MS} and Q'_o are chosen as primary parameters, the dependence on oxide thickness will be there, as seen from (2.2.6) and (4.5.32). Ideally, only process, temperature, and geometry information should have to be supplied to the model, and every effort should be extended to get as close to this goal as possible. The parameters should not be redundant; each should be responsible for one or more effects, which could not have been predicted without it. The parameter set should be chosen with parameter extraction in mind. In fact, parameter extraction should be kept in mind from the very beginning of model development.

3. The ideal physical model should be based on general expressions, derived from physics (not interpolation), which are valid in *all* regions of operation and are explicit functions of the terminal voltages.[†] Such a model remains an elusive goal at the time of this writing. Charge sheet models (Sec. 4.3) do make general expressions possible, which however are in terms of surface potentials (not terminal voltages). Very accurate evaluation of surface potential is needed, and this involves an implicit equation in principle (Sec. 4.3). Special attention must be paid to developing appropriate numerical techniques for such calculations. By using good approximations as initial “guesses,” surface potentials can be evaluated in only a few iterations. Several examples of CAD models based on the charge sheet approach can be found in the references.^{43,58,59,64,69} An alternative is to base the model on separate expressions for each region of inversion. This is sometimes preferred because expressions for individual regions of inversion can be simple (at least for weak and strong inversion), and give the current explicitly in terms of the terminal voltages; the evaluation of surface potentials is not required in this case. However, this approach has been responsible for numerous discontinuities in the derivatives of the current with respect to the bias voltages, which create problems with numerical convergence and cause serious errors in the prediction of small-signal parameters. To avoid such problems, all expressions should be combined into one single expression with the help of appropriate smoothing functions. This single expression can be used in all regions of inversion, but the smoothing functions are such that they “reveal” particular expressions in particular regions, while ensuring continuity at the boundaries between regions. Examples can be found in Ref. 44. A related approach has been used in the EKV model⁶¹ discussed in Sec. 4.8.
4. Whatever the approach used, it is imperative to ensure continuity of the current with respect to each terminal voltage, as well as *continuity of the derivatives* of the current with respect to each voltage. This is important for numerical convergence. A smooth variation of small-signal conductances will require continuity in the second-order derivatives of the current. However, *even higher-order derivatives must be continuous*, for numerical efficiency of the model and its parameter extraction, and for predicting accurately the nuances of the I - V characteristics; the latter is important, for example, in the evaluation of harmonic distortion.^{210–214} A graphical illustration⁵⁴ of how an otherwise accurate model can fail in predicting distortion is shown in Fig. 10.8. One should plan to ensure that the derivatives of any order are continuous, resulting in the so-called C_{∞} models.⁴⁴
5. It should be considered, right at the outset, whether the model will be source-referenced or body-referenced. The pros and cons of each approach have been discussed in Sec. 4.9. In principle, the two approaches should be equivalent, since a body-referenced model can be converted to a source-referenced model by

[†]Current-control alternatives have also been proposed.^{61,63,67}

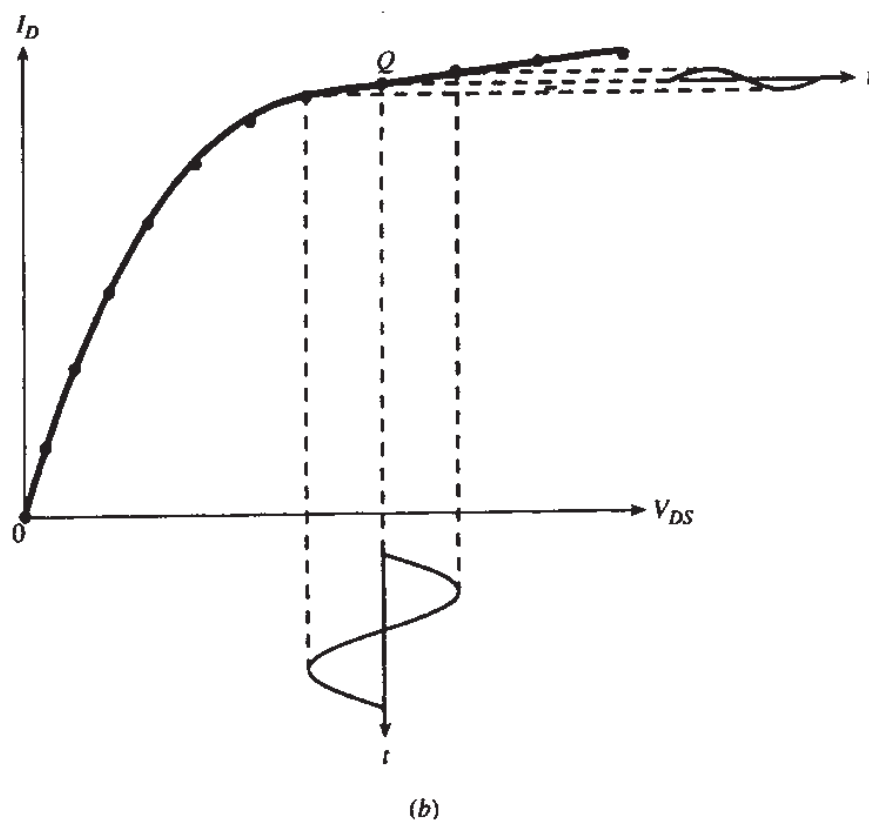
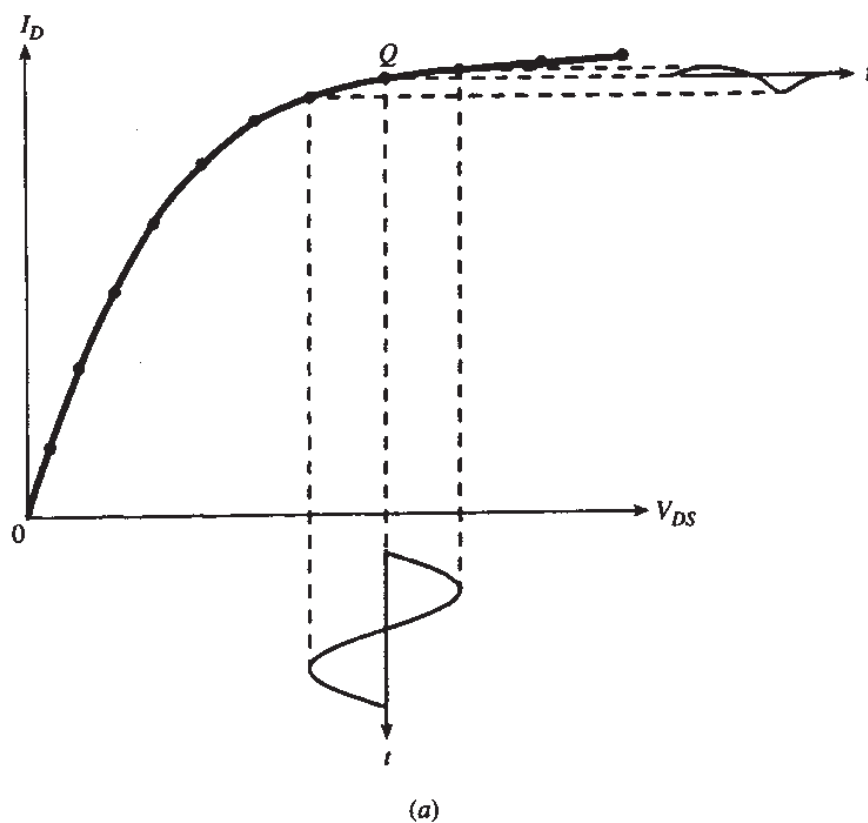


FIGURE 10.8

The I_D - V_{DS} curve as (a) measured and (b) modeled.⁵⁴ For a sinusoidal voltage variation, the shape of the corresponding current waveform in (b) is severely in error, although the error in modeling I_D is very small throughout (© 1994 by IEEE).

using $V_{GB} = V_{GS} + V_{SB}$ and $V_{DB} = V_{DS} + V_{SB}$. However, the practices adopted in most source-referenced models are responsible for a number of problems at $V_{DS} = 0$.⁷⁰ Among these are asymmetries of the source and drain capacitances at that point, and a discontinuous slope in g_{sd} versus V_{DS} plots (see Sec. 10.8). Given this history, if a source-referenced approach is to be chosen, the model developer must make sure that such problems do not occur. Body-referenced models do not exhibit such problems. On the other hand, the implementation of certain effects, such as velocity saturation, is easier in source-referenced models.

6. Many effects must be included in the model. This includes the various effects discussed in previous chapters of this book. It pays to spend as much time as possible to understand the physics of each such effect and to model it properly using physically meaningful parameters.
7. The modeling of charges must receive as much attention as the modeling of *dc* current. Most circuits do not operate at *dc*. During transients, or during steady-state *ac* operation, charging currents can be comparable to transport currents, and thus deserve equal attention. This makes necessary the modeling of charges not just in inversion, but also in depletion and accumulation. All charge expressions should be smooth functions of the terminal voltages.
8. The model should ideally be non-quasi-static.
9. If a non-quasi-static model is too complex to develop, then for large-signal transients a *complete* quasi-static model should be used (Chap. 7), which includes the effect of each terminal voltage on the charge passing through each of the terminals. This is one of the requirements for ensuring charge conservation.^{215–219}
10. The modeling of extrinsic effects (source, drain, gate, and substrate resistance, overlap and junction capacitances, junction leakage) must be done with an accuracy consistent with the accuracy expected of the intrinsic device model.
11. Remember: The modeling job is not finished until an efficient parameter extraction procedure has been developed to accompany it.

10.7.2 Considerations Related to Computer Implementation

The above considerations, together with the general requirements discussed in Sec. 10.6, dictate a number of requirements specific to the computer implementation of the model. Some of these are

1. There should be no IF statements in the code, in order to avoid the risk of discontinuities. Since it is very difficult to check continuity of every variable with respect to every variable, one cannot be sure that an IF statement will not cause a discontinuity or numerical problems. At times, the “no IF’s” rule may have to be bent in order to allow IF statements that turn off parts of the model not used for some particular parameter value combinations.⁷⁰ In no case, though, should an IF statement be encountered during a simulation run as a function of an electrical variable or of time. If necessary, smoothing functions can be used^{44,70} (see Example 10.1).

2. Although the model expressions are only intended to be accurate within certain ranges of voltages and currents, they must be well behaved even far outside those ranges.⁷⁰ In the course of numerical iterations, a simulator may try values well outside the intended ranges; if an exponential or a polynomial blows up, or a denominator becomes zero during such excursions, attempts to achieve numerical convergence can terminate unsuccessfully.
3. Charges should be used as the state variables for time-domain analysis. This, together with the use of complete charge models (see above) and the use of appropriate numerical techniques, will guarantee charge conservation,²¹⁵⁻²¹⁹ which is very important for simulating certain kinds of circuits.

In addition to these issues, others will have to be considered in particular cases. In general, at every step taken during model development, it pays to consider what that step implies in terms of computer implementation.

10.8 BENCHMARK TESTS†

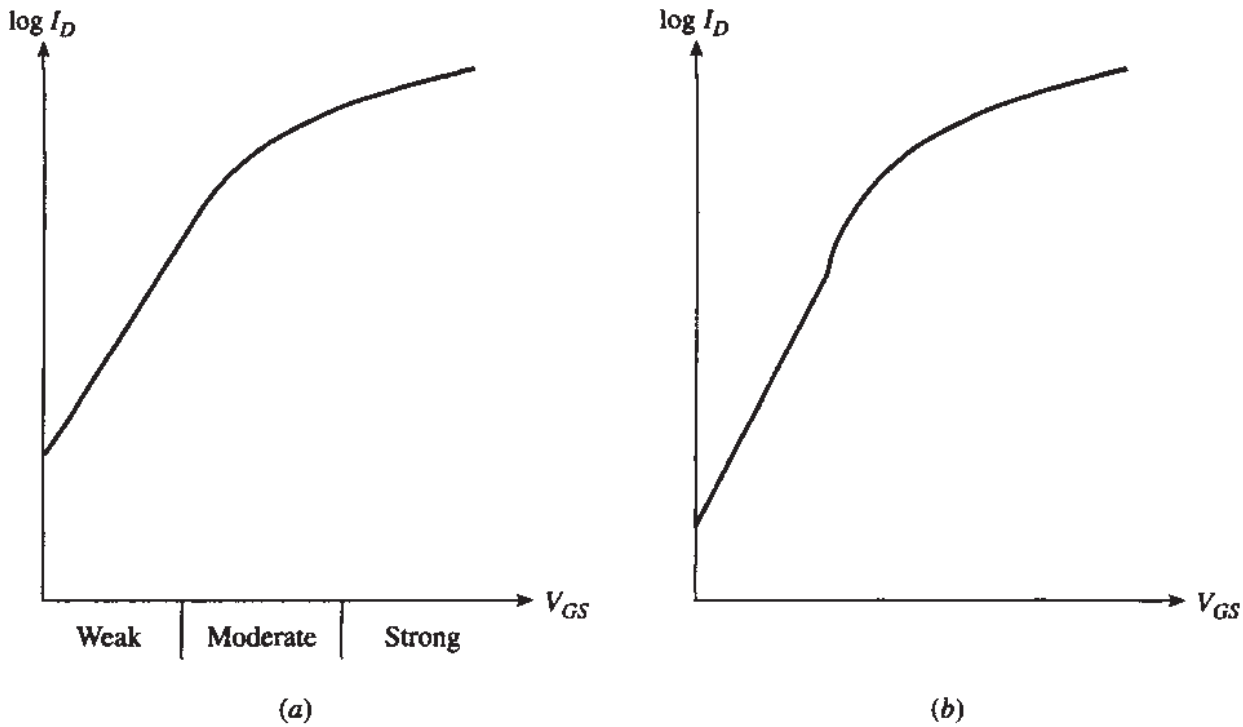
We now give several benchmark tests to evaluate MOSFET models. These tests have been found very useful over the years, as a necessary (but not sufficient) set of tests a model should pass before we can begin to trust it for analog work. Ideally, the tests should be *quantitative* comparisons to measured data; however, even if such data is not available, one can get very useful indications by running just the simulations indicated, since they will at least show whether the model being tested gives a correct *qualitative* behavior. Some examples of how well popular models fare when put to these tests will be given along the way, but we will avoid giving numerical comparisons, as we do not want to limit our comments to specific models with specific parameter values. All tests are to be done using nonminimum geometries (long and wide channels) at room temperature.

Test 1: Weak and moderate inversion current. For a V_{DS} value in the saturation region, plot I_D versus V_{GS} , with I_D on a *logarithmic* scale and including V_{GS} values well below threshold. The shape should be as illustrated by the solid line in Fig. 10.9a. Many popular models fail this test at least in the moderate inversion region, as shown in Fig. 10.9b.

Test 2: Transconductance-to-current ratio.

(a) **Continuity.** Plot the transconductance-to-current ratio, g_m/I_D (an important quantity for analog design) versus V_{GS} or versus $\log I_D$ (same range as for benchmark test 1, or at least a range including V_{GS} values around kinks such as the one in Fig.

†Unless indicated otherwise, these tests are from Y. Tsividis and K. Suyama, "MOSFET modeling for analog circuit CAD: Problems and prospects," *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 210–216, March 1994 (© 1994 by IEEE).

**FIGURE 10.9**

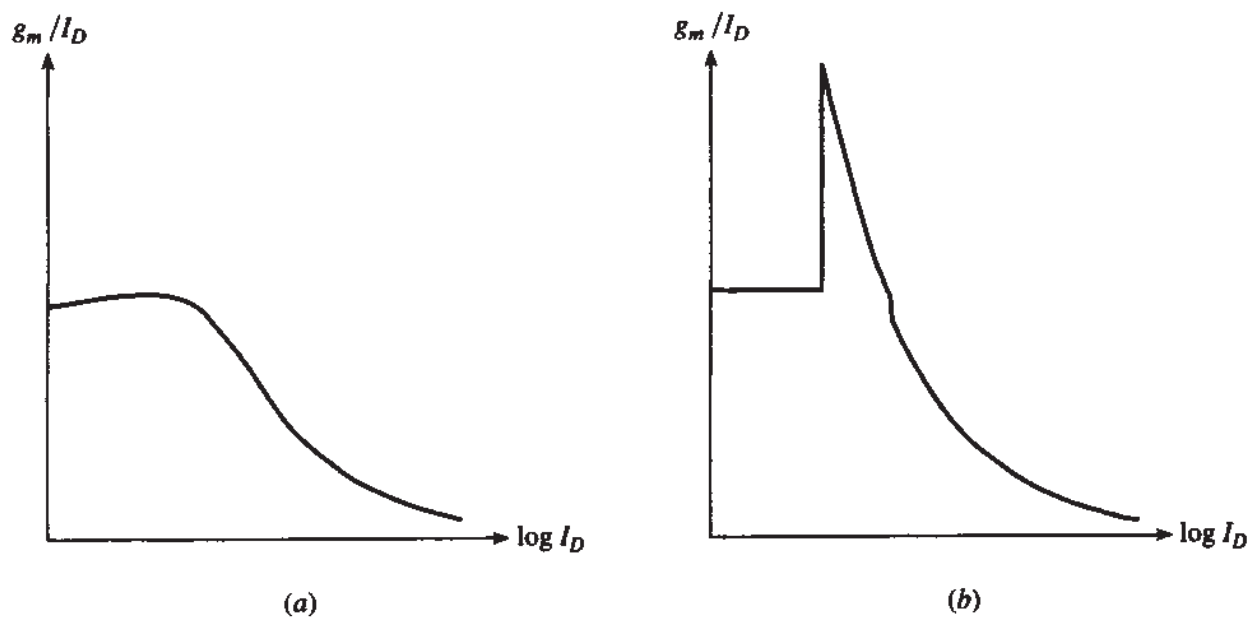
$\log I_D$ versus V_{GS} , for fixed V_{DS} and V_{SB} .⁵⁴ (a) A qualitatively correct curve. (b) Curve obtained using the level 2 Spice model (© 1994 by IEEE).

10.9b). Make sure that the V_{GS} (or I_D) spacing is fine enough, so that such kinks are not missed. Measurements give a shape as shown by the solid line in Fig. 10.10a.† Several models give results as shown in Fig. 10.10b. A large error occurs in the moderate inversion region.

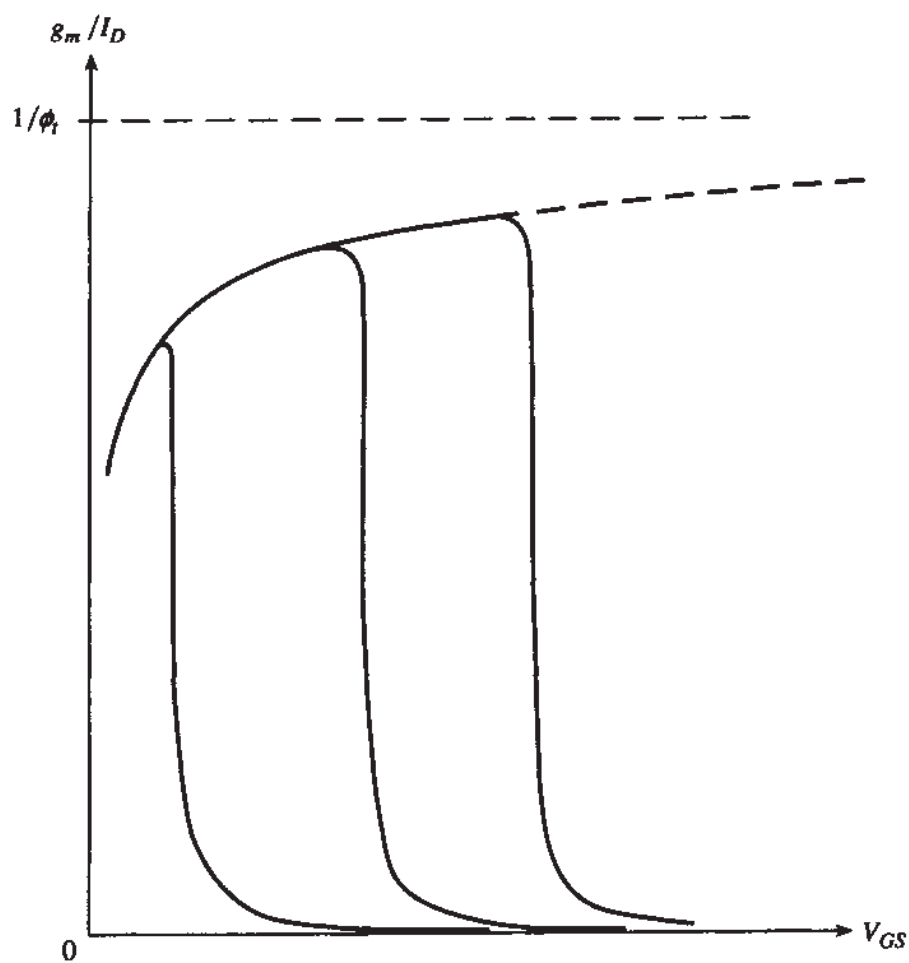
(b) Asymptotic behavior.²²⁰ Plot g_m/I_D versus V_{GS} , for several values of V_{SB} . The behavior should be as in Fig. 10.11. The top part of each curve corresponds to weak inversion. Using the accurate weak inversion model of (4.6.6)-(4.6.7), or the charge sheet model of Sec. 4.3.1, it can be found that this part tends asymptotically to $1/\phi_t$ as V_{SB} is increased, as shown in the figure. Some models predict a constant g_m/I_D in weak inversion, thus failing this test, which is referred to as the “Gummel tree-top test.”²²⁰⁻²²²

Test 3: Output conductance. Plot $g_o = dI_D/dV_{DS}$ versus V_{DS} , for a fixed V_{GS} value (or, better yet, obtain a family of such curves). The expected shape is as in Fig. 10.12a, but some models give a result as in Fig. 10.12b, depending on model parameter values. The reason is an unnatural transition from triode to saturation, shown in Fig. 10.12c (caused by an IF statement!). Other models do not produce such an

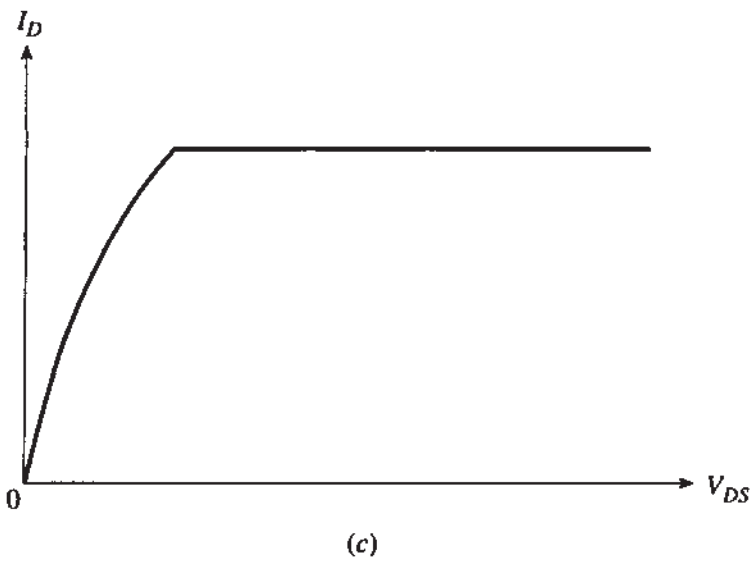
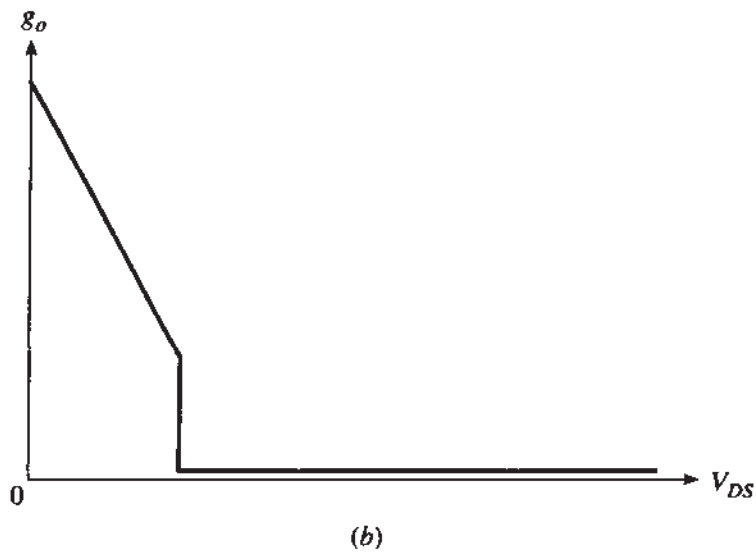
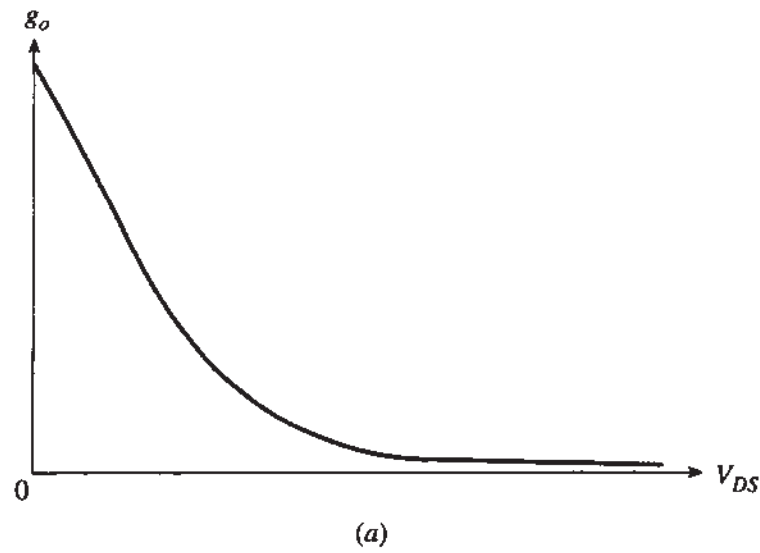
†The quantity g_m/I_D peaks in weak inversion, but does not become exactly constant in it. This is due to minute deviations from exponential behavior for the current (too minute to be noticeable in plots like the one of Fig. 9a), which are predicted by detailed charge-sheet models and by the detailed weak inversion model of (4.6.6)-(4.6.7).

**FIGURE 10.10**

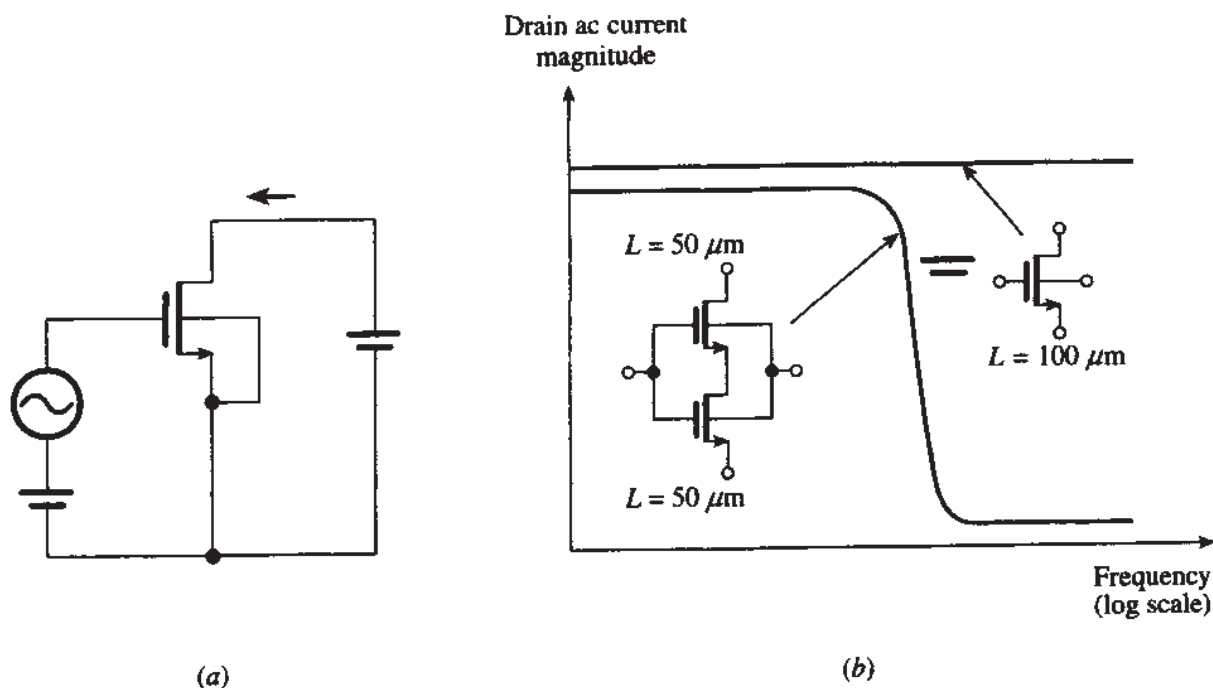
g_m/I_D versus $\log I_D$.⁵⁴ (a) A qualitatively correct curve. (b) A curve using the level 2 Spice model (© 1994 by IEEE).

**FIGURE 10.11**

g_m/I_D versus V_{GS} , for a fixed V_{DS} and for various V_{SB} values (Gummel tree-top test²²⁰⁻²²²).

**FIGURE 10.12**

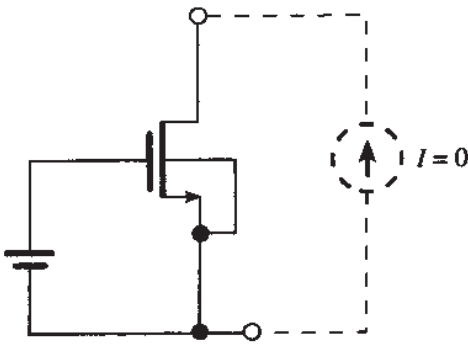
g_o versus V_{DS} for fixed V_{GS} and V_{SB} .⁵⁴ (a) A qualitatively correct curve; (b) a plot using a popular Spice model; (c) I_D versus V_{DS} for fixed V_{GS} and V_{SB} using the same model (© 1994 by IEEE).

**FIGURE 10.13**

(a) Circuit for obtaining ac response; (b) drain ac current magnitude versus frequency for a 100- μm -long MOSFET and an equivalent combination of two 50- μm -long devices, using the level 2 Spice model⁵⁴ (© 1994 by IEEE).

abrupt change but *still* predict g_o inaccurately in the transition from nonsaturation to saturation. Some models may also show discontinuity in plots of $\log g_o$ versus V_{GS} , with V_{GS} ranging from weak, through moderate, to strong inversion (and for several V_{SB} values). Discontinuities of g_o can also exist at $V_{DS} = 0$ (see test 7 below).

Test 4: High-frequency transadmittance. Take the simplest possible model (remove all parameters having to do with parasitics, such as junction and overlap capacitances, series resistances, etc.). In the device statement, make sure that the source and drain areas and perimeters are set to zero. We suggest the above simplifications to make clear what the following problem is caused by. Bias a 100- μm -long MOSFET in strong-inversion saturation, where the intrinsic gate-drain capacitance is zero. Use an ac source in series with the gate bias as shown in Fig. 10.13a, and obtain a frequency response for the ac drain current magnitude up to 10 GHz. Now break the device into two 50- μm -long devices, with their channels in series and with common gate and common substrate, and bias the combination as before. The combination should be equivalent to the single 100- μm device (remember, no junction area is supposed to exist at the intermediate point). Obtain the frequency response again. It should be the same as before. However, several common models give the behavior shown in Fig. 10.13b. The behavior is totally different at high frequencies. This is a result of the fact that the models used do not take into account non-quasi-static effects (Chap. 9); the behavior predicted for the 100- μm single device is, of

**FIGURE 10.14**

Circuit for simulating thermal noise in the triode region.⁵⁴ For this circuit, $V_{DS} = 0$.

course, totally unreasonable,[†] and contradicts both non-quasi-static models and measurements. The two-device combination does a better job at approximating reality, since it is a two-element lumped approximation of what is actually a distributed channel effect. (In fact, such combinations, with two or more elements, can be used in lack of non-quasi-static models, for high-frequency small-signal work; one should be careful, though, not to activate artificial short-channel effects in the subtransistors, and not to include extrinsic elements at intermediate channel points.) We note that non-quasi-static behavior has been experimentally demonstrated even in short-channel devices.²²³

Test 5: Thermal noise. Bias a device with a fixed V_{GS} in strong inversion, and at $V_{DS} = 0$ (a zero-value dc current source can be between drain and source, as shown in Fig. 10.14, if the simulator does not allow floating nodes). Run a noise simulation, for a frequency low enough so that the result is not affected by capacitances. Biased as indicated, the channel is equivalent to a resistor of value $R = 1/g_{sd}$, and should show a thermal noise voltage with power spectral density of $4kTR$ (e.g., $1.66 \times 10^{-16} \text{ V}^2/\text{Hz}$ for a g_{sd} of 10^{-4} A/V). Many models give, depending on implementation, a value that either is a couple of orders of magnitude too low or is even identically zero. The consequences are obvious for the design of analog circuits using MOSFETs as resistors.

Test 6: Flicker noise. Bias a device in strong-inversion saturation, and run a noise simulation at frequencies where $1/f$ noise should be dominant. The noise current can be converted to a voltage across a $1\text{-}\Omega$ resistor, placed in series with the drain (or, even better, a “noiseless resistor” implemented using a self-dependent voltage-controlled current source). Now, increase the channel area 10 times: Does the power spectral density of the equivalent input noise voltage (in V^2/Hz) decrease 10 times, as approximately observed in practice?

[†]In fact, the use of transcapacitors in some models can produce even worse errors, predicting that the ac current magnitude goes up with frequency (Fig. 9.25) (this effect will be seen only if it is not masked by extrinsic capacitance effects).

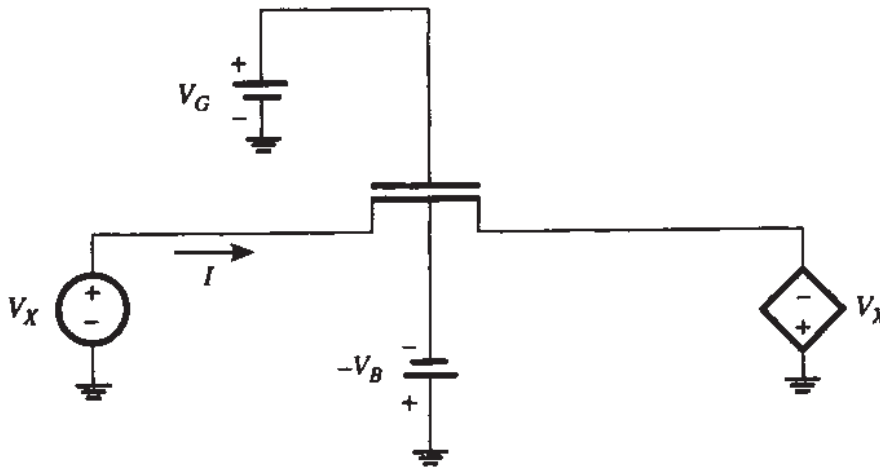


FIGURE 10.15
Circuit for tests 7 and 8.^{220-222,69}

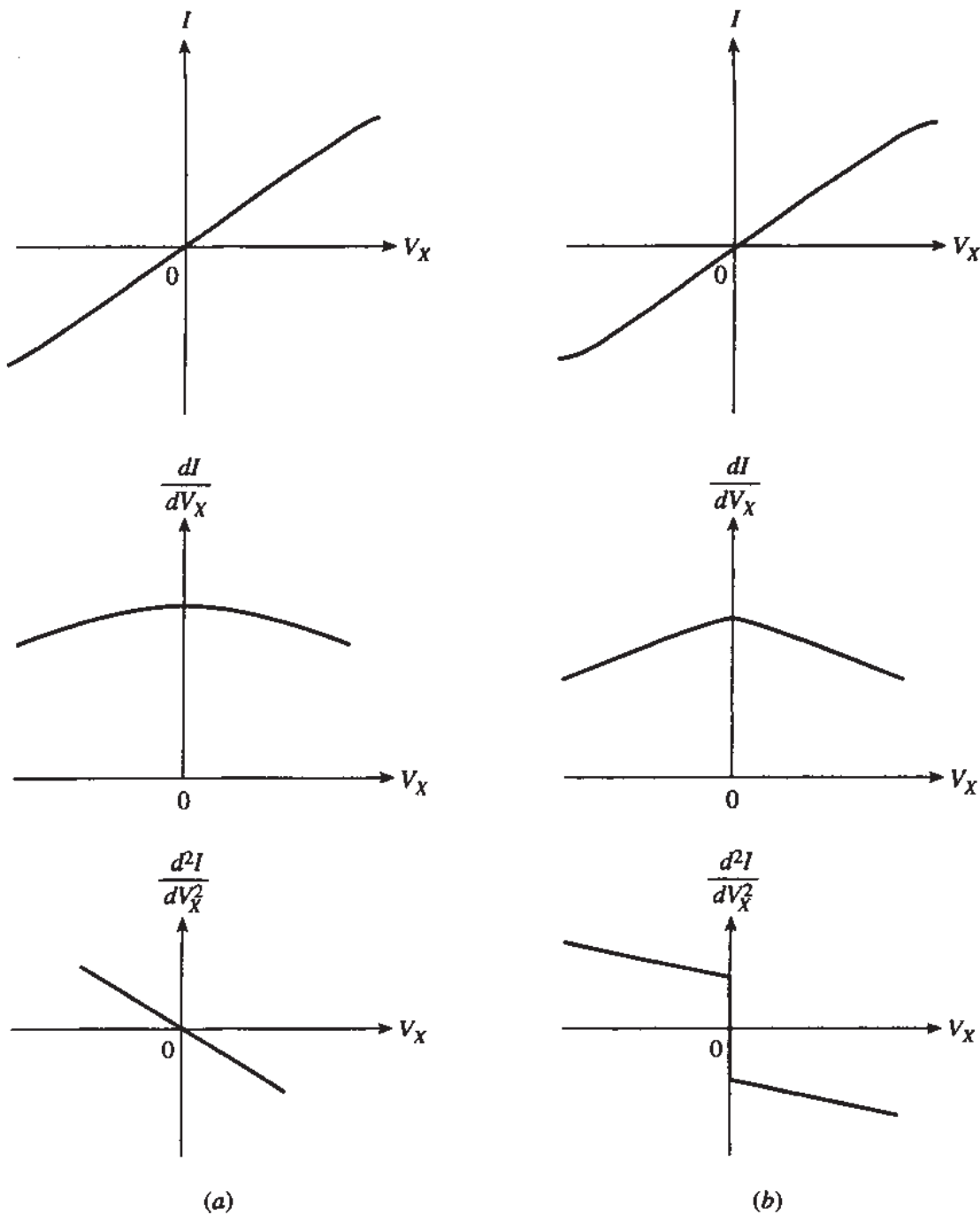
Test 7: Behavior of I_D at $V_{DS} = 0$.^{220-222,69} Bias a device in strong inversion, and apply a voltage $2V_X$ symmetrically between drain and source, as shown in Fig. 10.15. Plot I , dI/dV_X , and d^2I/dV_X^2 versus V_X . A correct model should exhibit the smooth behavior shown in Fig. 10.16a. Many models exhibit instead the behavior shown in Fig. 10.16b. This behavior can usually be traced to practices common in the implementation of source-referenced models.

Test 8: Charges and capacitances at $V_{DS} = 0$.^{220-222,69} Using again the circuit of Fig. 10.15, plot dQ/dV_X for the various charges in the device (Q_G , Q_B , Q_D , and Q_S in Chap. 7). These plots should be continuous at $V_X = 0$ ($V_{DS} = 0$). Instead, several models show a discontinuity at that point. In addition, a correct model should predict $C_{gs} = C_{gd}$ and $C_{bs} = C_{bd}$ at $V_{DS} = 0$, as shown in Fig. 10.17a; several models fail this test, as shown in Fig. 10.17b. Finally, discontinuities are found for the capacitances at the boundaries between different regions of operation, for example at $V_{GS} = V_{FB}$, at $V_{GS} = V_T$, and at V_{GS} corresponding to the onset of moderate or strong inversion. A test for such discontinuities is described elsewhere.²²⁴

Extra tests should be performed to make sure there are no discontinuities or kinks in any of the plots of I_D , $\log I_D$, and the small-signal parameters as a function of any of the terminal voltages, for a variety of device dimensions. And, of course, the model would have to be tested quantitatively, in comparison to measurements. Standardized ways to perform all these tests are described elsewhere.²²⁰⁻²²²

Many existing models will fail some or most of the above tests. And, of course, even if they pass some of them qualitatively, they still would have to pass them quantitatively, in comparison to measurements.

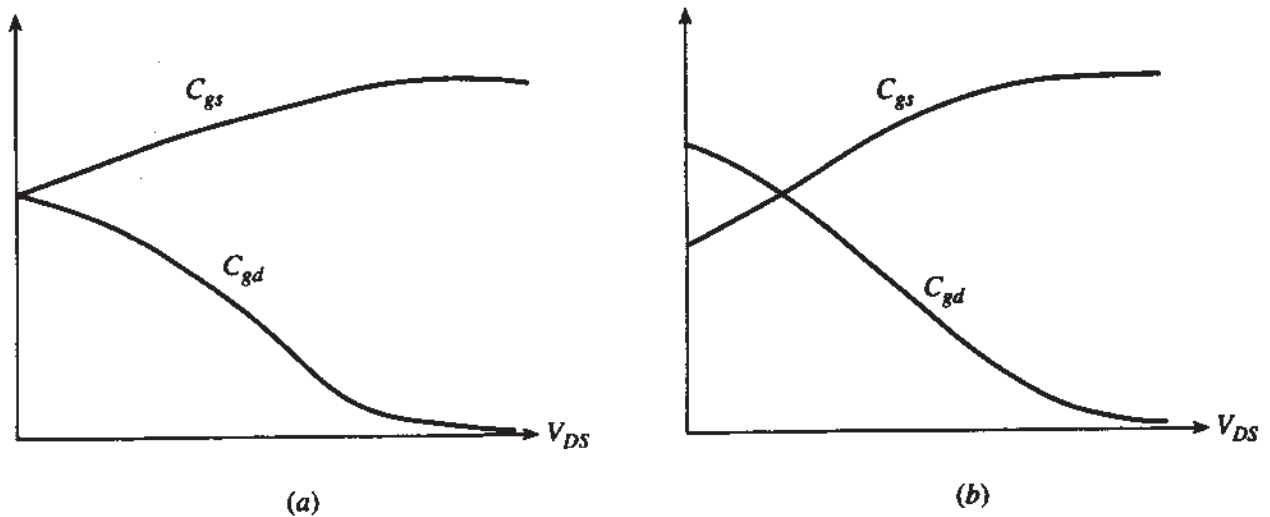
The above benchmark tests cover only some of the major problems which, at present, seem to be present in most popular CAD models. Other problem areas include the accuracy of capacitances and noise in the moderate inversion region, transient response under non-quasi-static conditions, the influence of the body effect along the channel on thermal noise, noise at frequencies where non-quasi-static effects are observed, noise of short-channel devices, effective mobility dependence on

**FIGURE 10.16**

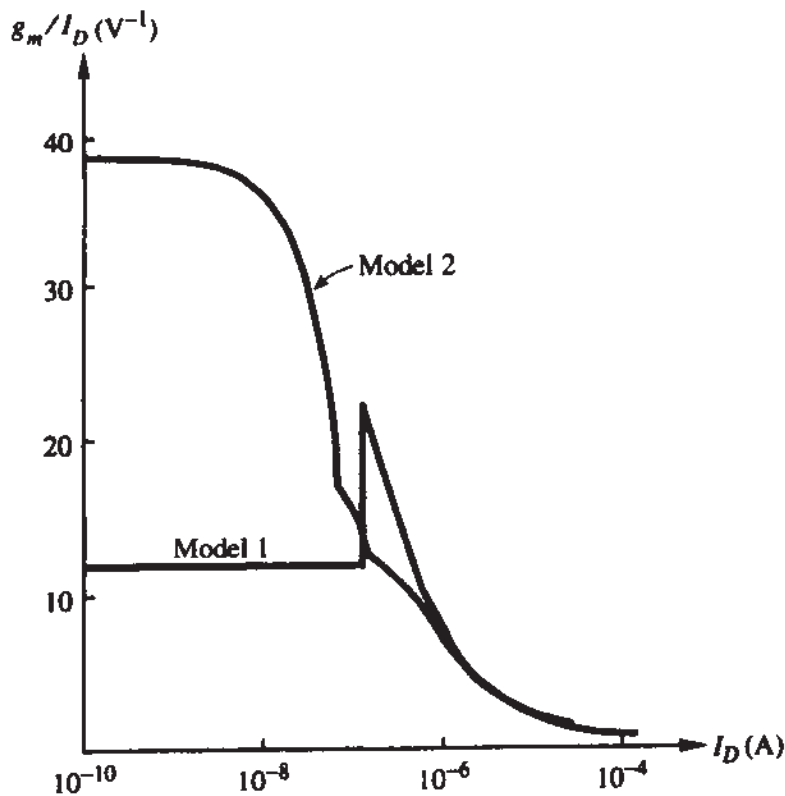
I , dI/dV_X , and d^2I/dV_X^2 versus V_X for the circuit of Fig. 10.15.^{220–222,69} (a) Qualitatively correct behavior; (b) common behavior of some popular models.

V_{SB} , etc. Most of these problems have been discussed in previous chapters. The problems mentioned as a rule get worse for devices with short and/or narrow channels. Also, particular models (and particular implementations of models in specific simulators) may have additional problems. Many circuit designers have formed a list of problems particular to the models/simulators they use.

Even if a model could in principle do a decent job over certain bias ranges, it often is not given the opportunity to do so because of poor parameter extraction. As an example,⁵⁴ the parameter sets provided by a well-known foundry service for two

**FIGURE 10.17**

C_{gs} and C_{gd} versus V_{DS} , for fixed V_{GS} and V_{SB} .⁶⁹ (a) Qualitatively correct behavior; (b) common behavior of some popular models.

**FIGURE 10.18**

g_m/I_D versus $\log I_D$ using two different models, with parameters as provided by a foundry service for the same fabrication process³⁴ (© 1994 by IEEE).

different models, for the *same* fabrication process (obtained from measurements on the *same* devices), were examined. I_D - V_{DS} curves in strong inversion (where the parameters were apparently extracted) gave relatively good agreement of one model with the other. However, a plot of g_m/I_D using the two models gave the results in Fig. 10.18. The figure speaks for itself (note that no comparison of accuracy of the two

models in weak inversion is implied here, since the foundry simply did not attempt to match the models to data in that region). Using the parameters from the same foundry also gave a 6-order-of-magnitude difference in $1/f$ noise predictions between the two models, a factor of 3 discrepancy in the saturation small-signal output conductance, etc. Such problems are due to extraction having only the digital designer in mind.

10.9 NONTECHNICAL CONSIDERATIONS

Anybody who is interested in developing a successful MOSFET model, which will improve upon existing ones and will be widely adopted, is considering a major undertaking. This is obvious from the many complicated effects that must be modeled correctly, and from the many considerations already listed above. But the job is not really finished when the equations have been written, or even when they have been implemented in a computer and tested. The job is not even finished when all results have been reported in conferences and journal publications. If the model developers desire to see their model widely adopted, there are a large number of other tasks to be completed. The model has to be implemented in a major circuit simulator such as Spice, and ideally in several popular versions of it. The interface between a model and a simulator has not been standardized yet, and this can be a source of delays and frustration. Detailed documentation must be written. Then the various CAD tool companies must be convinced to include the model in their products, which means that the source code must be provided to them. The various foundries must be convinced to support the model, and to extract parameters for it. It is a major commitment for a company to adopt a new model, and a lot of inertia is usually encountered. In fact, models with clear advantages have not been adopted for years because a commitment to other models, with all their problems, had already been made earlier on, a large amount of data had been accumulated using those models, and the models became entrenched. Finally, if the model gets adapted, its developers should commit themselves to answer questions, fix bugs, update the model, and in general support it.

Thus, while this author would like to encourage the development of new and better models, the readers who decide to take the challenge should be fully aware of the adventure they are getting into, especially if they want to see their model widely adopted.

REFERENCES

1. L. Nagel and R. Rohrer, "Computer analysis of nonlinear circuits, excluding radiation," *IEEE Journal of Solid-State Circuits*, vol. SC-6, pp. 166–182, 1971.
2. L. Nagel and D. O. Peterson, "Simulation program with integrated circuit emphasis," Electronics Research Laboratory Memorandum UCB/ERL M352, University of California, Berkeley, 1973.
3. W. T. Weeks, A. J. Jimenez, G. W. Mahoney, D. Mehta, H. Qassemzadeh, and T. R. Scott, "Algorithms for ASTAP—A network-analysis program," *IEEE Transactions on Circuit Theory*, vol. CT-20, pp. 628–634, 1973.
4. L. W. Nagel, "SPICE2: A computer program to simulate semiconductor circuits," Memorandum No. UCB/ERL-M520. Electronics Research Laboratory, University of California, Berkeley, May 1975.

5. L. O. Chua and P. M. Lin, *Computer-Aided Analysis of Electronic Circuits: Algorithms & Computational Techniques*, Prentice Hall, Englewood Cliffs, N.J., 1975.
6. R. Beresford and J. Domitrowich, "Survey of circuit simulators," *VLSI Design*, vol. 8, pp. 70–80, July 1987.
7. A. F. Schwarz, *Computer-Aided Design of Microelectronic Circuits and Systems*, vols. I and II, Academic Press, New York, 1987.
8. W. J. McCalla, *Fundamentals of CAD Simulation*, Kluwer Academic Publisher, Boston, 1988.
9. A. E. Ruehli (editor), *Circuit Analysis, Simulation and Design*, North-Holland, New York, 1986.
10. J. Vlach and V. Singhal, *Computer Methods for Circuit Analysis and Design*, Van Nostrand Reinhold, New York, 1993.
11. H. Shichman and D. A. Hodges, "Modeling and simulation of insulated-gate field-effect transistor switching circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-3, pp. 285–289, September 1968.
12. D. Frohman-Bentckowsky and L. Vadasz, "Computer-aided design and characterization of digital MOS integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-4, pp. 57–64, April 1969.
13. J. E. Meyer, "MOS models and circuit simulation," *RCA Review*, vol. 32, pp. 42–63, March 1971.
14. G. Merckel, J. Borel, and N. Z. Cupcea, "An accurate large-signal MOS transistor model for use in computer-aided design," *IEEE Transactions on Electron Devices*, vol. ED-19, pp. 681–690, May 1972.
15. T. Masuhara, J. Etoh, and M. Nagata, "A precise MOSFET model for low-voltage circuits," *IEEE Transactions on Electron Devices*, vol. ED-21, pp. 363–371, June 1974.
16. F. M. Klaassen, "A MOS model for computer-aided design," *Philips Research Reports*, vol. 31, pp. 71–83, 1976.
17. G. Merckel, "CAD models of MOSFETS," in *Process and Device Modelling for Integrated Circuit Design*, F. Van de Wiele, W. L. Engl, and P. G. Jespers (editors), Noordhoff, Leyden, The Netherlands, 1977.
18. F. M. Klaassen, "A MOST model for CAD with automated parameter determination," in *Process and Device Modelling for Integrated Circuit Design*, F. Van de Wiele, W. L. Engl, and P. G. Jespers (editors), Noordhoff, Leyden, The Netherlands, 1977.
19. F. M. Klaassen, "Review of physical models for MOS transistors," in *Process and Device Modelling for Integrated Circuit Design*, F. Van de Wiele, W. L. Engl, and P. G. Jespers (editors), Noordhoff, Leyden, The Netherlands, 1977.
20. A. Vladimirescu and S. Liu, "The Simulation of MOS ICs Using SPICE-2," University of California/Berkeley, Electronics Research Laboratory Document M80/7, 1980.
21. F. M. Klaassen and W. C. J. de Groot, "Modeling of scaled-down MOS transistors," *Solid-State Electronics*, vol. 23, pp. 765–772, 1980.
22. R. E. Oakley and R. J. Hocking, "CASMOS—an accurate MOS model with geometry-dependent parameters: I," *IEE Proceedings*, vol. 128, part I, pp. 239–247, December 1981.
23. H. I. Hanafi, L. H. Camnitz, and A. J. Dally, "An accurate and simple MOSFET model for computer-aided design," *IEEE Journal of Solid-State Circuits*, vol. SC-17, pp. 882–891, October 1982.
24. S. Liu and L. W. Nagel, "Small-signal MOSFET models for analog circuit design," *IEEE Journal of Solid-State Circuits*, vol. SC-17, pp. 983–998, December 1982.
25. B. Sheu, D. Scharfetter, and H. Poon, "Compact Short Channel IGFET Model (CSIM)," University of California/Berkeley, Electronics Research Laboratory Memorandum No. UCB/ERL M84/20, 1984.
26. A. L. Silburt, R. C. Foss, and W. F. Petrie, "An efficient MOS transistors model for computer-aided design," *IEEE Transactions on Computer-Aided Design*, vol. CAD-3, p. 104, January 1984.
27. B. Sheu, "MOS transistor modeling and characterization for circuit simulation," University of California/Berkeley, Electronics Research Laboratory Memorandum No. UCB/ERL M85/85, 1985.
28. G. T. Wright, "Simple and continuous MOSFET models for the computer-aided design of VLSI," *IEE Proceedings*, vol. 132, Part I, pp. 187–194, August 1985.
29. J. H. Satter, "The S-model: A highly accurate MOST model for CAD," *Solid State Electronics*, vol. 29, pp. 977–990, 1986.
30. S. L. Wong and C. A. T. Salama, "Improved simulation of p- and n-channel MOSFET's using an enhanced SPICE MOS3 model," *IEEE Transactions on Computer-Aided Design*, vol. CAD-6, pp. 586–591, July 1987.

31. B. J. Sheu, D. L. Scharfetter, P.-K. Ko, and M.-C. Jeng, "BSIM: Berkeley short-channel IGTET model for MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. SC-22, pp. 558–565, August 1987.
32. E. Sano and T. Kimura, "CMOS device model for accurate circuit simulation application," *Electronics and Communication in Japan*, Part 2, vol. 70-C, pp. 135–142, February 1987.
33. T.-T. Chia and G. J. Hu, "An accurate SPICE MOSFET model for digital and analog circuit simulations," *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 405–408, 1987.
34. S. Yu, A. F. Franz, T. G. Mihran, "A physical parametric transistor model for CMOS circuit simulation," *IEEE Transactions on Computer-Aided Design*, vol. 7, pp. 1038–1052, October 1988.
35. D. A. Divekar, "FET modeling for circuit simulation," Kluwer Academic Publisher, Boston, 1988.
36. S.-W. Lee and R. C. Rennick, "A compact IGFET model-ASIM," *IEEE Transactions on Computer-Aided Design*, vol. 7, pp. 952–975, September 1988.
37. J. A. Power and W. A. Lane, "Enhanced spice MOSFET model for analog applications including parameter extraction schemes," *Proceedings IEEE International Conference on Microelectronic Test Structures*, vol. 3, pp. 129–134, March 1990.
38. H. C. de Graaff and F. M. Klaassen, *Compact Transistor Modelling for Circuit Design*, Springer-Verlag, Vienna, 1990.
39. S. Veeraraghavan, "SSIM: A new charge-based MOSFET model," presented at MCNC Circuit Simulation Workshop, November 1990.
40. T. Pedron and G. Merckel, "ELSIMA: ELDO short-channel IGFET model for analog applications," *Proceedings of the European Solid State Device Research Conference*, Session 7A8, pp. 543–546, September 1990.
41. S. M. Gowda, B. J. Sheu, and J. S. Cable, "An accurate MOS transistor model for submicron VLSI circuits-BSIM_plus," *Proceedings IEEE Custom Integrated Circuits Conference*, pp. 23.2.1–23.2.4, San Diego, May 1991.
42. T. Sakurai and A. R. Newton, "A simple MOSFET model for circuit analysis," *IEEE Transactions on Electron Devices*, vol. 38, p. 887, April 1991.
43. A. R. Boothroyd, S. W. Tarasewicz, and C. Slaby, "MISNAN-A physically based continuous MOSFET model for CAD applications," *IEEE Transactions on Computer-Aided Design*, vol. 10, pp. 1512–1529, December 1991.
44. C. C. McAndrew, B. K. Bhattacharyya, and O. Wing, "A single-piece C_{∞} -continuous MOSFET model including subthreshold conduction," *IEEE Electron Device Letters*, vol. 12, pp. 565–567, October 1991.
45. M. Shur, T. A. Fjeldly, T. Ytterdal, and K. Lee, "Unified MOSFET model," *Solid-State Electronics*, vol. 35, pp. 1795–1802, December 1992.
46. J. A. Power, and W. A. Lane, "An enhanced SPICE MOSFET model suitable for analog applications," *IEEE Transactions on Computer-Aided Design*, vol. 11, pp. 1418–1425, November 1992.
47. N. Arora, *MOSFET Models for VLSI Circuit Simulation-Theory and Practice*, Springer-Verlag, Vienna, 1993.
48. Model Level 28, in *HSpice Users' Manual*, Meta-Software, Campbell, California, 1993.
49. P. Antognetti and G. Massobrio, *Semiconductor Device Modeling with Spice*, McGraw-Hill, New York, 1993.
50. K. Lee, M. Shur, T. A. Fjeldly, and T. Ytterdal, *Semiconductor Device Modeling for VLSI*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
51. R. M. D. A. Velghe, D. B. M. Klaassen, and F. M. Klaassen, "Compact MOS modeling for analog circuit simulation," *Digest International Electron Devices Meeting*, pp. 485–488, December 1993.
52. T. Skotnicki, C. Denat, P. Senn, G. Merckel, and B. Hennion, "A new analog/digital CAD model for sub-halfmicron MOSFETs," *Digest, International Electron Devices Meeting*, pp. 165–168, San Francisco, 1994.
53. S. M. Gowda and B. J. Sheu, "BSIM_plus: an advanced SPICE model for submicron MOS VLSI circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 13, pp. 1166–1170, September 1994.
54. Y. P. Tividis and K. Suyama, "MOSFET modeling for analog circuit CAD: problems and prospects," *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 210–216, March 1994.

55. N. D. Arora, R. Rios, C.-L. Huang, and K. Raol, "PCIM: a physically based continuous short-channel IGFET model for circuit simulation," *IEEE Transactions on Electron Devices*, vol. 41, pp. 988–997, June 1994.
56. R. M. D. A. Velghe, D. B. M. Klaassen, and F. M. Klaassen, "MOS model 9," Philips Electronics N. V., Unclassified Report NL-UR 003/94, 1994.
57. D. H. Cho and S. M. Kang, "A new deep submicron compact physical model for analog circuits," *Digest, 1994 IEEE Custom Integrated Circuits Conference*, pp. 41–44, May 1994.
58. B. Iñiguez and E. G. Moreno, "A physically based C_{∞} -continuous model for small-geometry MOSFET's," *IEEE Transactions on Electron Devices*, vol. 42, pp. 283–287, February 1995.
59. A. I. A. Cunha, M. C. Schneider, and C. G.-Montoro, "An explicit physical model for the long-channel MOS transistor including small-signal parameters," *Solid-State Electronics*, vol. 38, pp. 1945–1952, November 1995.
60. R. Rios, N. D. Arora, C.-L. Huang, N. Khalil, J. Faricelli, and L. Gruber, "A physical compact MOSFET model including quantum mechanical effects, for statistical circuit design applications," *Digest, International Electron Devices Meeting*, pp. 937–940, December 1995.
61. C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog Integrated Circuits and Signal Processing*, vol. 8, pp. 83–114, 1995.
62. A. Chatterjee, C. F. Machala, and P. Yang, "A submicron DC MOSFET model for simulation of analog circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, pp. 1193–1207, October 1995.
63. C. Enz, "The EKV model: A MOST model dedicated to low current and low voltage analogue circuit design and simulation," in *Low Power HF Microelectronics*, G. Machado (editor), IEE Circuits & Systems Series No. 8, IEE Book Publishing, London, 1996.
64. M. M.-Mattausch, U. Feldmann, A. Rahm, M. Bollu, and D. Savignac, "Unified complete MOSFET model for analysis of digital and analog circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, pp. 1–7, January 1996.
65. D. Foty, *MOSFET Modeling with SPICE-Principles and Practice*, Prentice-Hall PTR, Upper Saddle River, N.J., 1997.
66. Y. Cheng, M.-C. Jeng, Z. Liu, J. Huang, M. Chan, K. Chen, P. K. Ko, and C. Hu, "A physical and scalable I - V model in BSIM3v3 for analog/digital circuit simulation," *IEEE Transactions on Electron Devices*, vol. 44, pp. 277–287, February 1997.
67. A. I. A. Cunha, O. C. Gouveia-Filho, M. C. Schneider, and C. Galup-Montoro, "A current-based model for the MOS transistor," *Proceedings 1997 IEEE International Symposium on Circuits and Systems*, pp. 1608–1611, Hong Kong, June 1997.
68. S. H. Jen, B. J. Sheu, and Y. Oshima, "A unified approach to submicron DC MOS transistor modeling for low-voltage ICs," *Analog Integrated Circuits and Signal Processing*, vol. 12, pp. 107–118, 1997.
69. K. Joardar, K. K. Gullapalli, C. C. McAndrew, M. E. Burnham, and A. Wild, "An improved MOSFET model for circuit simulation," *IEEE Transactions on Electron Devices*, vol. 45, pp. 134–148, January 1998.
70. C. C. McAndrew, "Practical modeling for circuit simulation," *IEEE Journal of Solid-State Circuits*, vol. 33, pp. 439–448, March 1998.
71. B. R. Chawla, H. K. Gummel, and P. Kozak, "MOTIS—An MOS timing simulator," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 901–910, Dec. 1975.
72. T. Shima, T. Sugawara, S. Moriyama, and H. Yamada, "Three-dimensional table look-up MOSFET model for precise circuit simulation," *IEEE J. Solid-State Circuits*, vol. SC-17, pp. 449–453, June 1982.
73. T. Shima, H. Yamada, and R. Dang, "Table look-up MOSFET modeling system using a 2-D device simulator and monotonic piecewise cubic interpolation," *IEEE Trans. Computer-Aided Design*, vol. CAD-2, pp. 121–125, Apr. 1983.
74. J. L. Burns, A. R. Newton, and D. O. Pederson, "Active device table look-up models for circuit simulation," in *Proc. Int. Symp. Circuits Systems*, May 1983, pp. 250–253.

75. K. Sakui, T. Shima, Y. Hayashi, F. Horiguchi, and M. Ogura, "A simplified accurate three-dimensional table lookup MOSFET model for VLSI circuit simulation," in *Proc. IEEE Custom IC Conf.*, Portland, OR, 1985, pp. 347–351.
76. D. Divekar, D. Ryan, J. Chan, and J. Deutsch, "Fast and accurate table look-up MOSFET model for circuit simulation," in *Proc. IEEE Custom IC Conf.*, 1986, pp. 621–623.
77. P. Subramanian, "Modeling MOS VLSI circuits for transient analysis," *IEEE J. Solid-State Circuits*, vol. SC-21, pp. 276–285, Apr. 1986.
78. P. L. Meijer, "Table models for device modeling," in *Proc. of Int. Symp. Circuits Systems*, Helsinki, Finland, May 1988, pp. 2593–2596.
79. J. Barby, J. Vlach, and K. Singhal, "Polynomial splines for MOSFET model approximation," *IEEE Trans. Computer-Aided Design*, vol. 7, pp. 557–565, May 1988.
80. M. Conti, C. Turchetti, and G. Masetti, "A new methodology to produce accurate empirical models for VLSI MOSFETs," *Solid-State Electronics*, vol. 34, pp. 79–89, 1991.
81. M. G. Graham and J. J. Paulos, "Interpolation of MOSFET table data in width, length, and temperature," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 12, p. 1880, December 1993.
82. A. Rofougaran and A. A. Abidi, "A table lookup FET model for accurate analog circuit simulation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 12, pp. 324–335, February 1993.
83. M. G. Graham, J. J. Paulos, and D. W. Nychka, "Template-based MOSFET device model," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, p. 924, August 1995.
84. M. Conti, S. Orcioni, C. Turchetti, G. Soncini, and N. Zorzi, "Analytical device modeling for MOS analog IC's based on regularization and Bayesian estimation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, p. 1309, November 1996.
85. S. Selberherr, A. Schutz, and H. W. Pötzel, "MINIMOS—A two-dimensional MOS transistor analyzer," *IEEE Transactions on Electron Devices*, vol. ED-27, p. 1540, 1980.
86. W. L. Engl, H. K. Dirks, and B. Meinerzhagen, "Device Modeling," *Proceedings IEEE*, vol. 71, p. 10, 1983.
87. E. M. Buturla, P. E. Cottrell, B. M. Grossman, and K. A. Salsburg, "Finite-element analysis of semiconductor devices: The FIELDAY program," *IBM Journal of Research and Development*, vol. 25, pp. 131–146, 1981.
88. M. R. Pinto, C. S. Rafferty, and R. W. Dutton, "PISCES-II: Poisson and continuity equation solver," Stanford Electronic Laboratory Technical Report, September 1984.
89. T. Toyabe, H. Masuda, Y. Aoki, H. Shukuri, and T. Hagiwara, "Three-dimensional device simulator CADDETH with highly convergent matrix solution algorithm," *IEEE Transactions on Computer-Aided Design*, vol. CAD-4, pp. 482–488, 1985.
90. S. Selberherr, A. Schütz, and H. W. Pötzel, "MINIMOS—a two-dimensional MOS transistor analyzer," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 1540–1550, 1980.
91. R. W. Dutton and Z. Yu, *Technology CAD: Computer Simulation of IC Processes and Devices*, Kluwer, Boston, 1993.
92. P. Yang and P. Chatterjee, "Statistical modeling of small geometry MOSFET," *Digest, IEEE International Electron Devices Meeting*, pp. 286–289 (1982).
93. N. Herr, B. Garbs, and J. J. Barnes, "A statistical modeling approach for simulation of MOS VLSI circuit designs," *Digest, IEEE International Electron Devices Meeting*, p. 290 (paper 11.5), 1982.
94. P. Cox, et al., "Statistical modeling for efficient parametric yield estimation of MOS VLSI circuits," *IEEE Transactions on Electron Devices*, vol. ED-32, p. 471, 1985.
95. P. Cox, P. Yang, S. S. Mahant-Shetti, and P. Chatterjee, "Statistical modeling for efficient parametric yield estimation of MOS VLSI circuits," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 471–478, 1985.
96. C. J. B. Spanos and S. W. Director, "Parameter extraction for statistical IC process characterization," *IEEE Transactions on Computer-Aided Design*, vol. CAD-5, p. 66, 1986.

97. P. Yang, et al., "An integrated and efficient approach for MOS VLSI statistical circuit design, *IEEE Transactions on Computer-Aided Design*, vol. CAD-5, p. 5, 1986.
98. M. J. B. Bolt, A. Trip, and H. J. Verhagen, "Statistical Worst-Case MOS Parameter Extraction," *Proceedings on Microelectronic Test Structures*, Edinburgh, 1989.
99. J. A. Power, A. Mathewson, and W. A. Lane, "MOSFET statistical parameter extraction using multivariate statistics," *Proceedings IEEE International Conference on Microelectronic Test Structures*, vol. 4, pp. 209–214, March 1991.
100. J. A. Power, B. Donnellan, K. Burke, K. Moloney, A. Mathewson, and W. A. Lane, "Generation of MOS model parameters covering statistical process variations," *Proceedings of the 25th European Solid State Device Research Conference*, 1993.
101. J. A. Power, A. Mathewson, and W. A. Lane, "An approach for relating model parameter variabilities to process fluctuations," *Proceedings IEEE International Conference on Microelectronic Test Structures*, vol. 6, pp. 63–68, March 1993.
102. J. B. Shyu, G. Temes, and F. Krummenacher, "Random error effects in matched MOS capacitors and current sources," *IEEE Journal of Solid-State Circuits*, vol. SC-19, pp. 948–955, 1984.
103. S. Inohira, T. Shinmi, M. Nagata, T. Toyabe, and K. Iida, "A statistical model including parameter matching for analog integrated circuits simulation," *IEEE Transactions on Computer-Aided Design*, vol. CAD-4, pp. 621–628, 1985.
104. K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, "Characterization and modeling of mismatch in MOS transistors for precision analog design," *IEEE Journal of Solid-State Circuits*, vol. SC-21, pp. 1057–1066, 1986.
105. M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistor," *IEEE Journal of Solid-State Circuits*, vol. SC-24, pp. 1433–1439, 1989.
106. C. Michael and M. Ismail, "Statistical modeling of device mismatch for analog MOS integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-27, pp. 154–166, 1992.
107. D. E. Ward and K. Doganis, "Optimized extraction of MOS model parameters," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. CAD-1, p. 163, October 1982.
108. P. Yang and P. K. Chatterjee, "An optimal parameter extraction program for MOSFET models," *IEEE Transactions on Electron Devices*, vol. ED-30, pp. 1214–1219, September 1983.
109. K. Doganis and D. L. Scharfetter, "General optimization and extraction of IC device model parameters," *IEEE Transactions on Electron Devices*, vol. ED-30, pp. 1219–1228, September, 1983.
110. A. B. Bhattacharya, P. Ratnam, D. Nagchoudhuri, and S. C. Rustagi, "On-line extraction of model parameters of a long buried-channel MOSFET," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 545–550, 1983.
111. B. S. Messenger, "A fully automated MOS device characterization system for process-oriented integrated circuit design," Memorandum No. UCB/ERL M84/18, Electronic Research Laboratory, University of California, Berkeley, January 1984.
112. E. Khalily, P. H. Decher, and D. A. Teegarden, "TECAP2: An interactive device characterization and model development system," *Technical Digest, IEEE International Conference on Computer-Aided Design*, ICCAD-84, pp. 184–151, 1984.
113. K. L. Peng, S. Y. Oh, M. A. Fromowitz, and J. L. Moll, "Basic parameter measurement and channel broadening effect in the submicron MOSFET," *IEEE Electron Device Letters*, vol. EDL-5, pp. 473–475, 1984.
114. O. Melstrand, E. O'Neill, G. E. Sobelman, and D. Dokos, "A data base driven automated system for MOS device characterization, parameter optimization and modeling," *IEEE Transactions on Computer-Aided Design*, vol. CAD-3, pp. 47–51, January 1984.
115. P. Conway, C. Cahill, W. A. Lane, and S. U. Lidholm, "Extraction of MOSFET parameters using the simplex direct search optimization method," *IEEE Transactions on Computer-Aided Design*, vol. CAD-4, pp. 694–698, October 1985.
116. G. T. Wright and H. M. A. Gaffur, "Preprocessor modeling of parameter and geometry dependence of short and narrow MOSFET for VLSI circuit simulation, optimization, and statistics with SPICE," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 1240–1245, 1985.
117. C. Hao, B. Cabon-Till, S. Cristoloveanu, and G. Ghibaudo, "Experimental determination of short-channel MOSFET parameters," *Solid-State Electronics*, vol. 28, pp. 1025–1030, 1985.

118. K. Doganis and S. Hailey, "A unified physical device modeling environment," *IEEE 1986 Custom Integrated Circuit Conference*, pp. 203–207, 1986.
119. S. J. Wang, J. Y. Lee, and C. Y. Chang, "An efficient and reliable approach for semiconductor device parameter extraction," *IEEE Transactions on Computer-Aided Design*, vol. CAD-6, pp. 170–178, 1986.
120. W. Maes, K. M. De Meyer, and L. H. Dupas, "SIMPAP: A versatile technology independent parameter extraction program using new optimized fit strategy," *IEEE Transactions on Computer-Aided Design*, vol. CAD-5, pp. 320–325, 1986.
121. M. F. Hamer, "First-order parameter extraction on enhancement silicon MOS transistor," *IEE Proceedings*, vol. 133, Pt. I, pp. 49–54, 1986.
122. C. F. Machala, III, P. C. Pattnaik, and Ping Yang, "An efficient algorithm for the extraction of parameters with high confidence from nonlinear models," *IEEE Electron Device Letters*, vol. EDL-7, pp. 214–218, April 1986.
123. M. Sugimoto, "General-purpose model parameter extraction program with initial value exploration technique," *Technical Digest, IEEE Custom Integrated Circuits Conference, CICC-86*, pp. 624–627, 1986.
124. T. J. Krutsick, M. H. White, H.-S. Wong, and R. V. H. Booth, "An improved method of MOSFET modeling and parameter extraction," *IEEE Transactions on Electron Devices*, vol. ED-34, p. 1676, August 1987.
125. B. Lemaître and H. L. Zapf, "Optimization of drain current and small signal conductance of analytical MOS models for analog circuit simulation," *Proceedings of VLSI and Computers International Conference on Computer Technology, Systems and Applications*, pp. 341–344, 1987.
126. G. Ghibaudo, "New method for the extraction of MOSFET parameters," *Electronic Letters*, vol. 24, pp. 543–545, April 28, 1988.
127. B. Ankele, W. Holzl, and P. O'Levy, "Enhanced MOS parameter extraction and SPICE modeling," *Proceedings IEEE International Conference on Microelectronic Test Structures*, vol. 2, no. 1, pp. 73–78, March 1989.
128. L. Selmi, E. Sangiorgi, and B. Riccò, "Parameter extraction from *I-V* characteristics of single MOSFET's," *IEEE Transactions on Electron Devices*, vol. 36, p. 1094, June 1989.
129. C. Ciofi, M. Macucci, and B. Pellegrini, "A new measurement method of MOS transistor parameters," *Solid-State Electronics*, vol. 33, pp. 1065–1069, 1990.
130. C. C. McAndrew and P. A. Layman, "MOSFET effective length, threshold voltage, and series resistance determination by robust optimization," *IEEE Transactions on Electron Devices*, vol. 39, p. 2298, October 1992.
131. P. R. Karlsson and K. O. Jeppson, "An efficient parameter extraction algorithm for MOS transistor models," *IEEE Transactions on Electron Devices*, vol. 39, pp. 2070–2076, September 1992.
132. P. R. Karlsson and K. O. Jeppson, "Extraction of series-resistance-independent MOS transistor model parameters," *IEEE Electron Device Letters*, vol. 13, p. 581, November 1992.
133. M. Sharma and N. D. Arora, "OPTIMA: A Nonlinear Model Parameter Extraction Program with Statistical Confidence Region Algorithms," *IEEE Transactions on Computer-Aided Design*, vol. CAD-12, May 1993.
134. L. Selmi and B. Riccò, "Frequency-resolved measurements for the characterization of MOSFET parameters at low longitudinal field," *IEEE Transactions on Electron Devices*, vol. 42, p. 315, February 1995.
135. P. K. McLarty, S. Cristoloveanu, O. Faynot, V. Misra, J. R. Hauser, and J. J. Wortman, "A simple parameter extraction method for ultrathin oxide MOSFETs," *Solid-State Electronics*, vol. 38, pp. 1175–1177, June 1995.
136. G. A. S. Machado, C. C. Enz, and M. Bucher, "Estimating key parameters in the EKV MOST model for analogue design and simulation," *Proceedings ISCAS '95*, pp. 1588–1591, 1995.
137. M. Bucher, C. Lallement, and C. Enz, "An efficient parameter extraction methodology for the EKV most model," *Proceedings IEEE International Conference on Test Structures*, pp. 145–150, March 1996.
138. K. Terada and H. Muta, "A new method to determine effective MOSFET channel length," *Japanese Journal of Applied Physics*, vol. 18, pp. 953–959, 1979.

139. D. Takacs, W. Muller, and U. Schwabe, "Electrical measurement of feature sizes in MOS Si-gate VLSI technology," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 1368–1373, 1980.
140. J. G. J. Chern, P. Chang, R. F. Motta, and N. Godinho, "A new method to determine MOSFET channel length," *IEEE Electron Device Letters*, vol. EDL-1, pp. 170–173, 1980.
141. Y. R. Ma and K. L. Wang, "A new method to electrically determine effective MOSFET channel width," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 1825–1827, 1982.
142. K. L. Peng and M. A. Afromowitz, "An improved method to determine MOSFET channel length," *IEEE Electron Device Letters*, vol. EDL-3, pp. 360–362, 1982.
143. E. J. Korma, K. Visser, J. Snijder, and J. F. Verwey, "Fast determination of the effective channel length and the gate oxide thickness in polycrystalline silicon MOSFETs," *IEEE Electron Device Letters*, vol. EDL-5, pp. 368–370, 1984.
144. P. Vitanov, U. Schwabe, and I. Eisele, "Electrical characterization of feature sizes and parasitic capacitances using a single structure," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 96–100, 1984.
145. B. J. Sheu and P. K. Ko, "A simple method to determine channel widths for conventional and LDD MOSFETs," *IEEE Electron Device Letters*, vol. EDL-5, pp. 485–486, 1984.
146. B. J. Sheu and P. K. Ko, "A capacitance method to determine channel lengths for conventional and LDD MOSFETs," *IEEE Electron Device Letters*, vol. EDL-5, pp. 491–493, 1984.
147. S. E. Laux, "Accuracy of an effective channel length/external resistance extraction algorithm for MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 1245–1251, 1984.
148. M. R. Wordeman, J. Y.-C. Sun, and S. E. Laux, "Geometry effects in MOSFET channel length extraction algorithms," *IEEE Electron Device Letters*, vol. EDL-6, pp. 186–188, 1985.
149. J. Whitfield, "A modification on an improved method to determine MOSFET channel length," *IEEE Electron Device Letters*, vol. EDL-6, pp. 109–110, 1985.
150. C. T. Yao, I. A. Mack, and H. C. Lin, "Accuracy of effective channel-length extraction using the capacitance method," *IEEE Electron Device Letters*, vol. EDL-7, pp. 268–270, 1986.
151. L. Chang and J. Berg, "A derivative method to determine a MOSFETs effective channel length and width electrically," *IEEE Electron Device Letters*, vol. EDL-7, pp. 229–231, 1986.
152. J. Y.-C. Sun, M. R. Wordeman, and S. E. Laux, "On the accuracy of channel length characterization of LDD MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 1556–1562, 1986.
153. T. Y. Chan, A. T. Wu, P. K. Ko, and C. Hu, "A capacitance method to determine the gate-to-drain/source overlap length of MOSFET's," *IEEE Electron Device Letters*, vol. EDL-8, pp. 269–271, 1987.
154. G. J. Hu, C. Chang, and Y. T. Chia, "Gate-voltage-dependent effective channel length and series resistance of LDD MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 2469–2475, 1987.
155. J. Scarpulla, T. C. Mele, and J. P. Krusius, "Accurate criterion for MOSFET effective gate length extraction using the capacitance method," *Digest, IEEE International Electron Devices Meeting*, pp. 722–725, 1987.
156. J. Scarpulla and J. P. Krusius, "Improved statistical method for extraction of MOSFET effective channel length and resistance," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 1354–1359, 1987.
157. J. H. Satter, "Effective length and width of MOSFET's determined with three transistors," *Solid-State Electronics*, vol. 30, pp. 821–828, 1987.
158. D. J. Mountain, "Application of electrical effective channel length and external resistance measurement techniques to a submicrometer CMOS process," *IEEE Transactions on Electron Devices*, vol. 36, pp. 2499–2505, November 1989.
159. W. M. Coughran, Jr., W. Fichtner, and E. Grosse, "Extracting transistor charges from device simulations by gradient fitting," *IEEE Transactions on Computer-Aided Design*, vol. 8, p. 380, April 1989.
160. D. J. Mountain, "Application of electrical effective channel length and external resistance measurement techniques to a submicrometer CMOS process," *IEEE Transactions on Electron Devices*, vol. ED-36, pp. 2499–2505, 1989.

161. J. Ida, A. Kita, and F. Ichikawa, "Accurate characterization of gate-N⁻ overlapped LDD with the new Leff extraction method," *Digest, IEEE International Electron Devices Meeting*, pp. 219–222, 1990.
162. K. K. Ng and J. R. Brews, "Measuring the effective channel length of MOSFETs," *IEEE Circuits and Devices Magazine*, vol. 6, pp. 33–38, November 1990.
163. N. D. Arora, L. A. Bair, and L. M. Richardson, "A new method to determine the MOSFET effective channel width," *IEEE Transactions on Electron Devices*, vol. 37, pp. 811–814, March 1990.
164. Y.-T. Chia and G. J. Hu, "A method to extract gate-bias-dependent MOSFET's effective channel width," *IEEE Transactions on Electron Devices*, vol. 38, p. 424, February 1991.
165. C. Duvvury, J. L. Wise, C. F. Machala, and P. Yang, "A novel method to determine gate-drain overlap in sub-micron transistors," *Digest, International Electron Devices Meeting*, pp. 489–492, December 1993.
166. C. C. McAndrew, P. A. Layman, and R. A. Ashton, "MOSFET effective channel width determination by nonlinear optimization," *Solid-State Electronics*, vol. 36, pp. 1717–1723, 1993.
167. J.-C. Guo, S.-S. Chung, and C. C.-H. Hsu, "A new approach to determine the effective channel length and the drain-and-source series resistance of miniaturized MOSFET's," *IEEE Transactions on Electron Devices*, vol. 41, p. 1811, October 1994.
168. H.-H. Li and C.-Y. Wu, "A novel extraction technique for the effective channel length of MOSFET devices," *IEEE Transactions on Electron Devices*, vol. 42, p. 856, May 1995.
169. Y.-S. Jean and C.-Y. Wu, "A new extraction algorithm for the metallurgical channel length of conventional and LDD MOSFET's," *IEEE Transactions on Electron Devices*, vol. 43, p. 946, June 1996.
170. C.-L. Huang, J. V. Faricelli, D. A. Antoniadis, N. A. Khalil, and R. A. Rios, "An accurate gate length extraction method for sub-quarter micron MOSFET's," *IEEE Transactions on Electron Devices*, vol. 43, p. 958, June 1996.
171. Z. Latif, A. O.-Conde, J. J. Liou, and F. J. G. Sanchez, "A study of the validity of capacitance-based method for extracting the effective channel length of MOSFET's," *IEEE Transactions on Electron Devices*, vol. 44, pp. 340–343, February 1997.
172. J. J. Paulos, D. A. Antoniadis, and Y. P. Tsividis, "Measurement of intrinsic capacitances of MOS transistors," *Technical Digest, IEEE International Solid-State Circuits Conference*, San Francisco, pp. 238–239, February 1982.
173. H. Iwai and S. Kohyama, "On-chip capacitance measurement circuits in VLSI structures," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 1622–1626, 1982.
174. J. Oristian, H. Iwai, J. Walker, and R. Dutton, "Small geometry MOS transistor capacitance measurements method using simple on-chip circuit," *IEEE Electron Device Letters*, vol. EDL-5, pp. 395–397, 1984.
175. C. T. Yao and H. C. Lin, "Comments on 'Small geometry MOS transistor capacitance measurements method using simple on-chip circuit'," *IEEE Electron Device Letters*, vol. EDL-6, p. 63, 1985.
176. J. Oristian, H. Iwai, J. Walker, and R. Dutton, "A reply to comments on 'Small geometry MOS transistor capacitance measurements method using simple on-chip circuit'," *IEEE Electron Device Letters*, vol. EDL-6, pp. 64–67, 1985.
177. J. J. Paulos and D. A. Antoniadis, "Measurement of minimum geometry MOS transistor capacitances," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 357–363, 1985.
178. H. Iwai, J. Oristian, J. Walker, and R. Dutton, "A scaleable technique for the measurements of intrinsic MOS capacitance with atto-Farad range," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 344–356, 1985.
179. H. Ishiuchi, Y. Matsumoto, S. Sawada, and O. Ozawa, "Measurement of intrinsic capacitance of lightly doped drain (LDD) MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 2238–2242, 1985.
180. K. C. K. Weng and P. Yang, "A direct measurement technique for small geometry MOS transistor capacitances," *IEEE Electron Device Letters*, vol. EDL-6, pp. 40–42, 1985.
181. M. Furukawa, H. Hatano, and K. Hanihara, "Precision measurement technique of integrated MOS capacitor mismatching using a simple on-chip circuit," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 938–944, 1986.

182. P. Leclaire, "High resolution intrinsic MOS capacitance measurement system," EESDERC 1987, *Technical Digest*, pp. 702, 1987.
183. B. J. Sheu and P. K. Ko, "Measurement and modeling of short-channel MOS transistor gate capacitances," *IEEE Journal of Solid-State Circuits*, vol. SC-22, pp. 464-472, 1987.
184. Y. T. Yeow, "Measurement and numerical modeling of short channel MOSFET gate capacitance," *IEEE Transactions on Electron Devices*, vol. ED-35, pp. 2510-2519, 1987.
185. N. D. Arora, D. A. Bell, and L. A. Bair, "An accurate method of determining MOSFET gate overlap capacitance," *Solid-State Electronics*, vol. 35, pp. 1817-1822, 1992.
186. P. P. Suciú and R. L. Johnston, "Experimental derivation of the source and drain resistance of MOS transistors," *IEEE Transactions on Electron Devices*, vol. ED-27, pp. 1556-1162, 1980.
187. M. H. Seavey, "Source and drain resistance determination for MOSFETs," *IEEE Electron Device Letters*, vol. EDL-5, pp. 479-481, 1984.
188. B. J. Sheu, C. Hu, P. K. Ko, and F.-C. Hsu, "Source-and-drain series resistance of LDD MOS-FETs," *IEEE Electron Device Letters*, vol. EDL-5, pp. 365-367, 1984.
189. K. K. Ng and W. T. Lynch, "Analysis of the gate-voltage dependence series resistance of MOS-FETs," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 965-972, 1986.
190. C. Y. Hwang, T.-C. Kuo, and J. C. S. Woo, "Extraction of gate dependent source/drain resistance and effective channel length in MOS devices at 77 K," *IEEE Transactions on Electron Devices*, vol. 42, p. 1863, October 1995.
191. H. G. Lee, S. Y. Oh, and G. Fuller, "A simple and accurate method to measure the threshold voltage of an enhancement-mode MOSFET," *IEEE Transactions on Electron Devices*, vol. ED-29, pp. 346-348, 1982.
192. S. Jain, "Measurement of threshold voltage and channel length of submicron MOSFETs," *Proceedings IEE*, Pt. I, vol. 135, pp. 162-164, 1988.
193. M. J. Deen and Z. X. Yan, "A new method for measuring the threshold voltage of small-geometry MOSFETs from subthreshold conduction," *Solid-State Electronics*, vol. 33, pp. 503-511, 1990.
194. F. Corsi, C. Marzocca, and G. V. Portacci, "New experimental technique for fast and accurate MOSFET threshold extraction," *Electronics Letters*, vol. 29, pp. 1358-1360, July 1993.
195. F. H. De La Moneda, H. N. Kotecha, and M. Shatzkes, "Measurement of MOSFET constant," *IEEE Electron Device Letters*, vol. EDL-3, pp. 10-12, 1982.
196. P.-M. D. Chow and K.-L. Wang, "A new AC technique for accurate determination of channel charge and mobility in very thin gate MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 1299-1304, 1986.
197. A. Hairapetian, D. Gitlin, and C. R. Viswanathan, "Low-temperature mobility measurements on CMOS devices," *IEEE Transactions on Electron Devices*, vol. ED-36, pp. 1448-1445, 1989.
198. G. Sh. Gildenblat, C.-L. Huang, and N. D. Arora, "Split C-V measurements of low temperature MOSFET inversion layer mobility," *Cryogenics*, vol. 29, pp. 1163-1166, 1989.
199. C. L. Huang, J. Faricelli, and N. D. Arora, "A new technique for measuring MOSFET inversion layer mobility," *IEEE Transactions on Electron Devices*, vol. ED-40, pp. 1134-1139, 1993.
200. T. Y. Chan, P. K. Ko, and C. Hu, "A simple method to characterize substrate current in MOS-FETs," *IEEE Electron Device Letters*, vol. EDL-5, pp. 505-507, 1984.
201. D. Lau, G. Gildenblat, C. G. Sodini, and D. E. Nelsen, "Low temperature substrate current characterization of n-channel MOSFETs," *Digest, IEEE International Electron Devices Meeting*, pp. 565-568, 1985.
202. R. V. H. Booth and M. H. White, "An experimental method for determination of the saturation point of a MOSFET," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 247-251, 1984.
203. W. Y. Jang, C. Y. Wu, and H. J. Wu, "A new experimental method to determine the saturation voltage of a small-geometry MOSFET," *Solid-State Electronics*, vol. 31, pp. 1421-1431, 1988.
204. G. Dambrine, A. Cappy, F. Heliodore, and E. Playez, "A new method for determining the FET small-signal equivalent circuit," *IEEE Trans. Microwave Theory Tech.*, vol. 36, no. 7, pp. 1151-1159, July 1988.
205. D. Lovelace, J. Costa, and N. Camilleri, "Extracting small-signal model parameters of silicon MOSFET transistors," in *IEEE MTT-S International Microwave Symposium Digest*, pp. 865-866, 1994.

206. H. Ikeda, "An elegant method for measuring MOST drain-source conductance in the saturation current region," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-21, pp. 234–236, August 1972.
207. N. R. Draper and H. Smith, *Applied Regression Analysis*, Wiley, New York, 1981.
208. Y. Tividis and G. Masetti, "Problems in precision modeling of the MOS transistor for analog applications," *IEEE Transactions on Computer-Aided Design*, vol. CAD-3, pp. 72–79, January 1983.
209. J. L. D'Arcy and R. C. Rennick, "MOSFET parameter optimization for accurate output conductance modeling," *Proceedings 1985 IEEE Custom Integrated Circuits Conference*, Portland, Ore., pp. 512–515.
210. G. Groenewold and W. J. Lubbers, "Systematic distortion analysis for MOSFET integrators with use of a new MOSFET model," *IEEE Transactions on Circuits and Systems—II*, vol. 41, pp. 569–580, September 1994.
211. E. A. M. Klumperink, C. H. J. Mensik, and P. M. Stroet, "Comment on low-voltage CMOS transconductance cell based on parallel operation of triode and saturation transconductors," *Electronics Letters*, vol. 30, pp. 1824–1825, October 27, 1994.
212. R. van Langevelde and F. M. Klaassen, "Effect of gate-field dependent mobility degradation on distortion analysis in MOSFET," *IEEE Transactions on Electron Devices*, vol. ED-44, no. 11, pp. 2044–2052, 1997.
213. R. van Langevelde and F. M. Klaassen, "Accurate drain conductance modeling for distortion analysis in MOSFETs," *Digest, International Electron Devices Meeting*, 1997.
214. P. Wambacq and W. Sansen, *Distortion Analysis of Analog Integrated Circuits*, Kluwer, Dordrecht, The Netherlands, 1998.
215. P. Yang, B. Epler, and P. Chatterjee, "An Investigation of the Charge Conservation Problem for MOSFET Circuit Simulation," *IEEE Journal of Solid State Circuits*, vol. SC-18, pp. 128–138, 1983.
216. J. G. Fossum, H. Jeong, and S. Veeraraghavan, "Significance of the channel-charge partition in the transient MOSFET model," *IEEE Transactions on Electron Devices*, vol. ED-33, pp. 1621–1623, October 1986.
217. M. F. Sevat, "On the channel charge division in MOSFET modeling," *Digest of IEEE International Conference on Computer-Aided Design*, pp. 204–207, 1987.
218. M. A. Cirit, "The Meyer model revisited: why is charge not conserved?" *IEEE Transactions on Computer-Aided Design*, vol. CAD-8, pp. 1033–1037, 1989.
219. I. W. Smith, H. Statz, H. A. Haus, and R. A. Pucel, "On charge nonconservation in FET's," *IEEE Transactions on Electron Devices*, vol. ED-34, pp. 2565–2568.
220. *Report of Sematech Compact Modeling Workshop*, Sematech, Sunnyvale, Calif., 1995.
221. "Benchmarks for compact MOSFET models," *Electronic Industries Alliance*, <http://www.eia.org/eig/cmc>.
222. "IEEE recommended practices P1485 on test procedures for microelectronic MOSFET circuit simulator model validation," <http://ray.eeel.nist.gov/modval/database/contents/reports/micromosfet/standard.html>.
223. R. Singh, A. Juge, R. Joby, and G. Morin, "An investigation into the nonquasistatic effects in MOS devices with a wafer S-parameter technique," *Proceedings IEEE International Conference on Microelectronic Test Structures*, Barcelona, March 1993.
224. U. Feldmann, A. Rahm, and M. Miura-Mattausch, "Benchmarking MOS transistor models with respect to capacitances and charges for analog applications," *Proceedings IEEE Symposium on Circuits and Systems*, vol. 2, pp. 1352–1355, 1995.

PROBLEMS

- 10.1. In Sec. 10.4, we discussed a method for obtaining the value of V_T using measured data in the nonsaturation region. Show that, under certain conditions, V_T can also be determined using data in the saturation region, by plotting $\sqrt{I_D}$ vs. V_{GS} . [Hint: See (4.5.37)]. What problems do you see with this technique, for practical devices?

- 10.2. Show that a possible procedure for extracting graphically the values of γ , ϕ_0 , and V_{FB} in (4.5.32) from measured data is the following. (a) Plot V_T vs. $\sqrt{V_{SB} + \phi_0}$, using an initial guess for the value of ϕ_0 . (b) If the plot turns out to be not a straight line, modify the value of ϕ_0 and try again, until approximately a straight line is obtained. Explain how, from the final plot, you can obtain the values of the above parameters.
- 10.3. Propose a method for obtaining the value of α in the model of (4.5.37), from measured data.
- 10.4. Devise a procedure for extracting the value of ΔW in (10.4.2) from measurements, for a transistor with nonnegligible source and drain resistance.
- 10.5. Shown in Fig. P10.1 are measured characteristics for an n -channel transistor, giving I_D versus V_{DS} with V_{GS} a parameter; each set of curves is for a different value of V_{SB} . Choose a *single* value for each of the quantities μ , V_{T0} , γ , α , and ϕ_0 , so that the approximate model equations are reasonably matched to the measurements. Assume $W = 1 \mu\text{m}$, $L = 1.4 \mu\text{m}$, $C'_{ox} = 3 \times 10^{-7} \text{ F/cm}^2$, a constant mobility, and no short-channel effects.

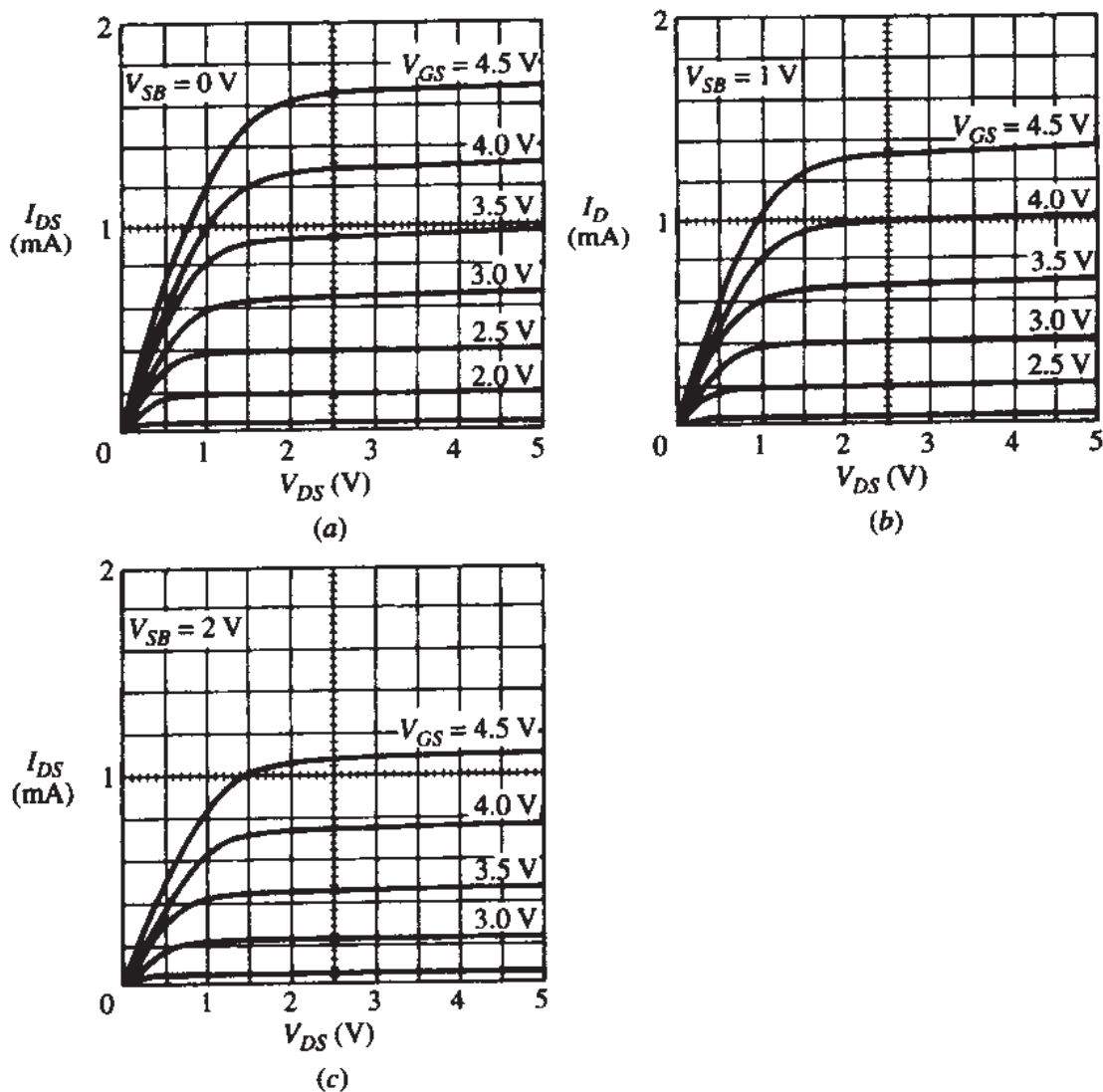


FIGURE P10.1

- 10.6.** Propose a method for extracting from measurements the value of \mathcal{E}_c in (10.3.2), assuming for simplicity that V_T is independent of V_{DS} .
- 10.7.** Implement the complete charge sheet model of Sec. 4.3.1 on the computer. You will need a numerical routine for calculating the surface potentials ψ_{so} and ψ_{SL} , as explained in the above section. [In lack of actual measurements, “data” produced using this model will be assumed to be “measurements” for some of the problems that follow.] Assume $V_{FB} = -0.9$ V, $N_A = 10^{17}$ cm⁻³, $t_{ox} = 100$ Å, and $W = L = 2$ μm.
- 10.8.** Determine μ , V_{T0} , γ , ϕ_0 , and α in the model of (4.5.37), so that this model matches “measurements” produced as in Prob. 10.7, for V_{GS} between 1.2 V and 3 V, V_{DS} between 0 and 3.3 V, and V_{SB} between 0 and 2 V. Assume that C'_{ox} has been determined from measurements to be 3.45×10^{-7} F/cm².
- 10.9.** Using the “measurements” produced as in Prob. 10.7, extract the value of V_T (a) as explained in Sec. 10.4, (b) as discussed in Prob. 10.1, (c) assuming V_T is the value of V_{GS} at which $I_D = (W/L) \times 0.1$ μA, and (d) using (4.5.32) with $\phi_0 = 2\phi_F$. The above four quantities are often assumed to be one and the same. Do you agree? If not, assume that $\phi_0 = 2\phi_F + \Delta\phi$, as in (2.5.23); what value should be used for $\Delta\phi$, in order for (4.5.32) to predict the value in (a)?
- 10.10.** (a) Devise a technique for extracting the values of μ_0 and θ in (4.10.20). Assume θ_B is negligible.
 (b) Use your technique to determine μ_0 and θ if it is given that, for a transistor with $V_T = 0.6$ V, V_{GS} values of 1.5, 2.0, 2.5, and 3.0 V produce measured μ values of 563, 532, 501, and 475 cm²/(V·s), respectively.
- 10.11.** Run the benchmark tests of Sec. 10.8 on one or more popular CAD models used at your institution and comment.

APPENDIX A

ENERGY BANDS AND RELATED CONCEPTS

A.1 ENERGY BANDS

The energy band model is a rather involved one in solid-state physics. Here we only summarize some of its basic features for interested readers. More details can be found in the references provided in Chap. 1. The following material can be understood in the context of Sec. 1.2.

For semiconductors, the energy band model can be illustrated as shown in Fig. A.1. The horizontal axis corresponds to distance in the semiconductor, whereas the vertical axis corresponds to electron energy. In an *intrinsic* semiconductor, electrons bound to their parent atom have energy no larger than E_v ; they are said to “be in the valence band.” An electron with a total energy of at least E_c becomes liberated from the parent atom and is said to “be in the conduction band.” Such an electron leaves “behind” a hole in the valence band. The energy of holes is measured in a direction opposite from that of electrons because of their opposite charge (i.e., hole energy increases downward in Fig. A.1). If an electron acquires a total energy $E > E_c$ (e.g., because of thermal vibration of the lattice), the difference $E - E_c$ corresponds to net kinetic energy as the electron moves in the crystal lattice. E_c itself represents the potential energy of the free electron. Energy levels between E_v and E_c are not occupied in the intrinsic semiconductor under discussion. Such energies belong to the so-called “forbidden band gap,” which for silicon has a width $E_g = E_c - E_v$ of 1.12 eV at

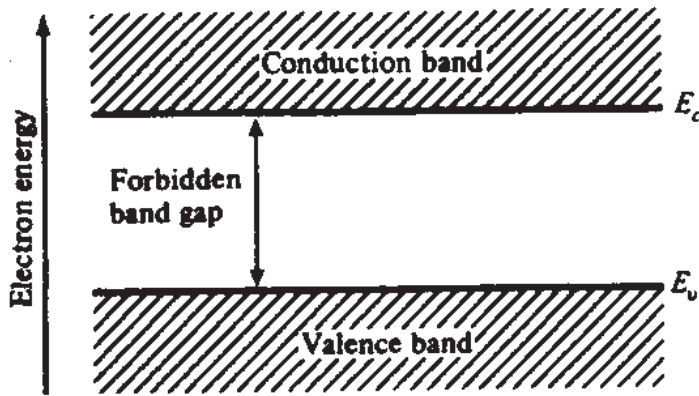


FIGURE A.1
Energy band model for a semiconductor.

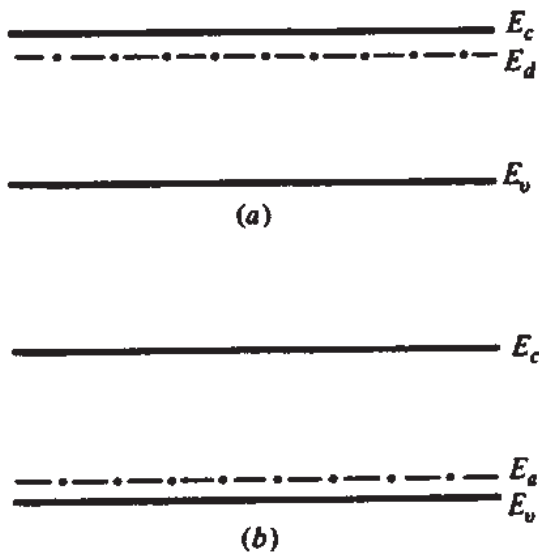


FIGURE A.2
Indicating energy levels corresponding to impurity atoms in (a) *n*-type extrinsic semiconductors (E_d), (b) *p*-type extrinsic semiconductors (E_a).

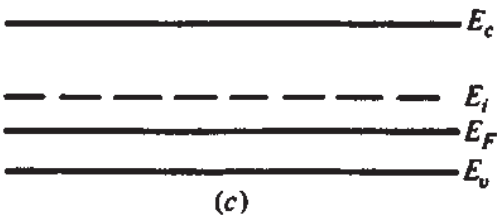
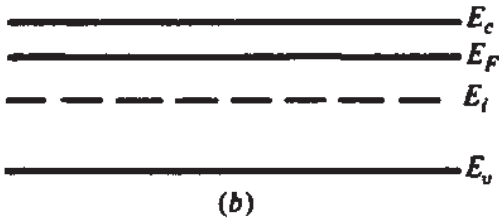
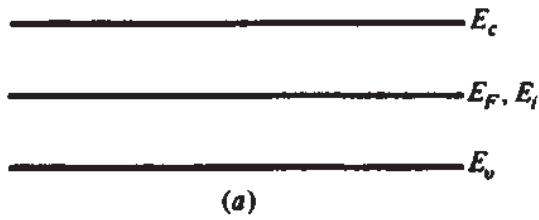
300 K, slightly decreasing with temperature. At room temperature, few electrons acquire that much energy in an intrinsic semiconductor and, therefore, few are liberated. Thus the intrinsic carrier concentration n_i is small.

In an *extrinsic n*-type semiconductor, the “extra” electron of a donor atom (Sec. 1.2) corresponds to an energy level E_d only slightly below E_c (Fig. A.2a). Then at room temperature practically all such electrons (one per donor atom) have enough thermally acquired energy to enter the conduction band (i.e., to be set free). Similarly, in a *p*-type semiconductor, the vacancy in an acceptor atom corresponds to an energy level E_a only slightly above E_v (Fig. A.2b), and thus it is easy for an electron from the valence band to fill the vacancy and leave “behind” a hole.

Although electrons and holes are governed by Fermi-Dirac statistics, these statistics practically reduce to Maxwell-Boltzmann statistics in nondegenerate semiconductors. According to these statistics, the electron and hole concentrations at equilibrium can be expressed as follows:

$$n = n_i e^{(E_F - E_i)/(kT)} \quad (\text{A.1})$$

$$p = n_i e^{(E_i - E_F)/(kT)} \quad (\text{A.2})$$

**FIGURE A.3**

Relative position of intrinsic energy level (E_i) and Fermi energy (E_F) for (a) intrinsic, (b) n -type, and (c) p -type semiconductors.

where n_i is the intrinsic carrier concentration, k is Boltzmann's constant, T is the absolute temperature, E_i is the "intrinsic energy level" (located very close to the middle of the forbidden energy gap), and E_F is the Fermi energy. For intrinsic semiconductors $E_F = E_i$ (Fig. A.3a). For n -type extrinsic semiconductors $E_F > E_i$ (Fig. A.3b), and for p -type extrinsic semiconductors $E_F < E_i$ (Fig. A.3c). If n or p for a nondegenerate semiconductor is known, $E_F - E_i$ can be calculated from (A.1) or (A.2). For the semiconductor material to be nondegenerate, the resulting E_F must not be too close to either end of the forbidden energy gap. The relation $E_v + 3kT < E_F < E_c - 3kT$ should be satisfied for such a material.

In equilibrium, E_F is constant throughout a semiconductor. To accommodate this, the levels E_c , E_i , and E_v may have to "bend" accordingly. This is illustrated in Fig. A.4b, showing the energy band diagram for a pn junction (Fig. A.4a) in equilibrium (no external bias) (Sec. 1.5). Similarly, for a junction of p -type silicon to intrinsic silicon (Fig. A.5a) in equilibrium (again, no external bias) the energy band diagram is as in Fig. A.5b. In both cases, the difference between any two among E_c , E_i , and E_v is kept constant with distance. For any point along the horizontal axis in Figs. A.4 and A.5, the distance between E_F and E_i , and their relative position, must be such as to give the correct values of n and p when (A.1) and (A.2) are used. The actual values of n and p can be found with the help of Poisson's equation (Appendix B).

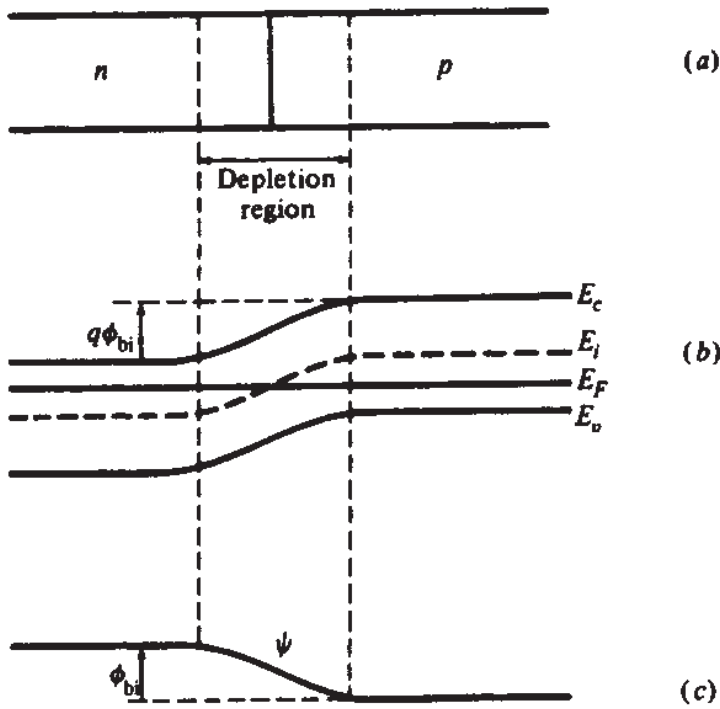


FIGURE A.4

A pn junction with related energy band diagram and potential in equilibrium.

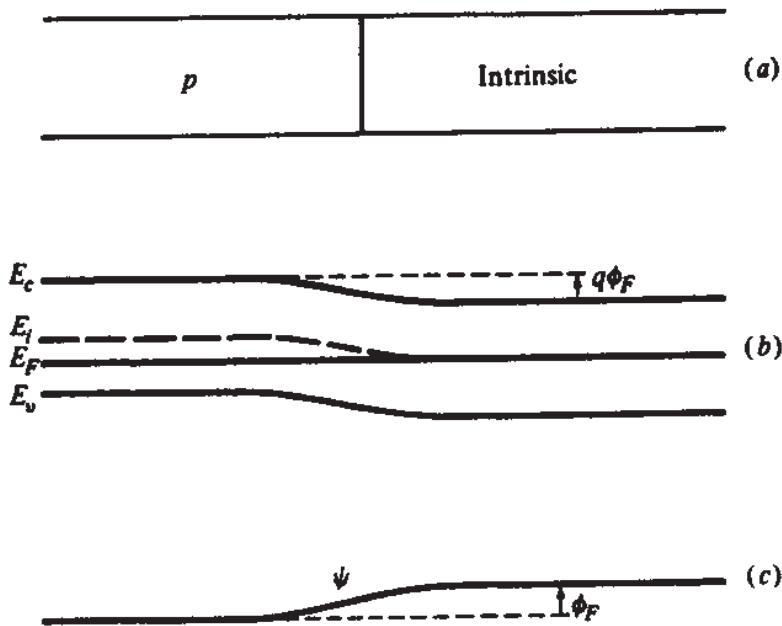


FIGURE A.5

A junction between p-type and intrinsic silicon with related energy band diagram and potential in equilibrium.

Let now $\Delta E_c = \Delta E_i$ be the electron potential energy difference between two points 1 and 2 in a semiconductor at equilibrium. Then the electrostatic potential difference $\Delta\psi$ between them (potential energy difference per unit charge) will be

$$\Delta\psi = \frac{\Delta E_i}{-q} \quad (\text{A.3})$$

Thus, electrostatic potential varies in a direction opposite from E_i , E_c , or E_v , as illustrated in Figs. A.4c and A.5c.

Using (A.1), the electron concentrations at the two points considered above will be $n_1 = n_i \exp [(E_F - E_{i1})/kT]$ and $n_2 = n_i \exp [(E_F - E_{i2})/kT]$. Dividing these and using (A.3) gives

$$\frac{n_1}{n_2} = e^{\phi_{12}/\phi_i} \quad (\text{A.4})$$

with $\phi_i = kT/q$; (A.4) is the same as (1.2.7). Similarly, one is led to (1.2.11) starting from (A.2).

In energy band treatments, the Fermi potential ϕ_F is defined for the cases of Fig. A.3 as follows†:

$$\phi_F = \frac{E_F - E_i}{-q} \quad (\text{A.5})$$

Our use of the Fermi potential in Sec. 1.4 (Fig. 1.11) is consistent with the above. This can be easily checked by relating Fig. A.5 to Fig. 1.11 and using (A.3) appropriately. Also, it is easy to check directly on the energy band diagram of the pn junction in Fig. A.4b that the contact potential of the n side to the p side (built-in potential, ϕ_{bi}) is given by (1.5.1).

If the doping concentration becomes too high, E_F becomes too close to the conduction or valence band (say, within $3kT$) and then the semiconductor is said to be degenerate. Then (A.1) and (A.2) do not hold. Thus, ϕ_F cannot be found from (1.4.2) or (1.4.3) in such cases. For a degenerate semiconductor with $E_F \approx E_c$ (n type) or $E_F \approx E_v$ (p type), we have $|E_F - E_i| \approx E_g/2$, where E_g is the band gap energy. For silicon at room temperature, this means that ϕ_F from (A.5) is approximately -0.56 V for n type and $+0.56$ V for p type.

A.2 CONTACT POTENTIALS AND WORK FUNCTIONS

In energy band treatments, contact potentials (Sec. 1.4) are handled using the so-called work functions.‡ Consider as an example n -type and p -type materials, initially separated as in Fig. A.6. Here E_R represents the so-called *vacuum energy level*, corresponding to the energy of an electron when it is removed from the material so that it is not influenced by it. The difference between E_R and a Fermi energy E_F is called *work function* and is denoted by W . It is a measure of “how difficult” it is for

†In some treatments, no minus sign is used in the denominator of the fraction in (A.5).

‡The treatment provided here is not rigorous. A careful treatment should be based on a foundation of thermodynamics.

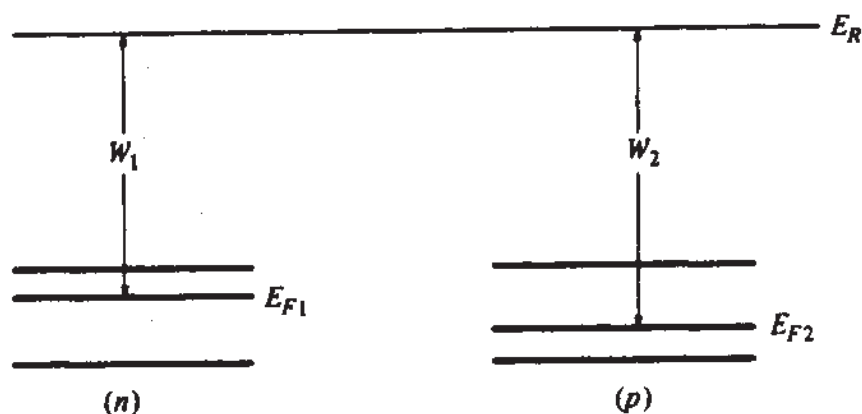


FIGURE A.6

Energy band diagrams for n and p materials separated from each other, and corresponding work functions.

an electron to leave its host material. In Fig. A.6, we have $W_1 < W_2$, and thus electrons find it easier to leave the n -type material. This causes the following effect. When the two materials are brought together to form a junction as in Fig. A.4, initially electrons diffuse from the n side to the p side (and holes diffuse in the opposite direction), as discussed in Sec. 1.5. With negative charges thus increasing on the p side, eventually the potential of that region with respect to the n side becomes so negative as to inhibit a further net tendency for diffusion. This happens when the average energy of the electrons on the p side (relative to that on the n side) has increased over its value before the materials were brought together, by the amount $W_2 - W_1$. This "causes the two Fermi levels to line up," as shown in Fig. A.4, and the other energy levels to increase from left to right by the amount $W_2 - W_1$, corresponding to a drop in electrostatic potential of $(W_2 - W_1)/q$; this is the contact potential of the n side to the p side. The same idea is used to handle contact potentials between dissimilar materials, e.g., a metal and a semiconductor. In general, then, the contact potential ϕ_{J_1, J_2} of a material J_1 to a material J_2 is given by

$$\phi_{J_1, J_2} = \frac{W_{J_2} - W_{J_1}}{q} \quad (\text{A.6})$$

where W_{J_1} and W_{J_2} are the corresponding work functions. As noted in Sec. 1.4, work functions are not easy to measure and, in fact, have to be modified for use in MOS device modeling.

A.3 QUASI-FERMI LEVELS AND CURRENTS

The discussion so far in this appendix has been limited to equilibrium conditions (Sec. 1.2). In the absence of equilibrium (A.1) and (A.2) cannot be used, and n and p must be found from other considerations (Refs. 1–12 in Chap. 1). The system cannot be characterized by a Fermi level which is constant throughout as in equilibrium. In

some treatments it is found convenient to introduce two quantities E_{Fn} and E_{Fp} , called *quasi-Fermi energy levels* (or *imrefs*) for electrons and holes, respectively. These are often defined as the quantities which satisfy the relations (see, e.g., the references in Chap. 1)†

$$n = n_i e^{(E_{Fn} - E_i)/(kT)} \quad (\text{A.7})$$

$$p = n_i e^{(E_i - E_{Fp})/(kT)} \quad (\text{A.8})$$

where n and p are the correct electron and hole concentration values. As is obvious by comparing the above equations to (A.1) and (A.2), in equilibrium we can set $E_{Fn} = E_{Fp} = E_F$, and thus $np = n_i^2$. Otherwise, $E_{Fn} \neq E_{Fp}$ and then $np \neq n_i^2$. This situation is encountered, for example, in the depletion region of a pn junction with external bias applied.

The quantities E_{Fn} and E_{Fp} can be related to electron current and hole current, respectively, in a simple manner. Thus, consider one-dimensional current flow in the x direction, uniformly distributed throughout a cross-sectional area A .‡ Let $\psi(x)$ and $n(x)$ be the electrostatic potential and the electron concentration at x . We can express the total current due to electrons $I_n(x)$, caused by both drift and diffusion, by using (1.3.11a; in footnote) and (1.3.17). If the current is defined in the *positive* x direction (*opposite* from that in Fig. 1.5 or 1.8), we will have

$$I_n(x) = qA \left[-\mu_n n(x) \frac{d\psi}{dx} + D_n \frac{dn}{dx} \right] \quad (\text{A.9})$$

where, μ_n and D_n are the electron mobility and the diffusion constant, respectively. As follows from (A.3), if $\psi(x)$ varies with x so will $E_i(x)$. Dividing both sides of that equation by Δx and letting Δx approach zero, we obtain:

$$\frac{d\psi}{dx} = -\frac{1}{q} \frac{dE_i}{dx} \quad (\text{A.10})$$

From (A.7) we obtain

$$\frac{dn}{dx} = \frac{1}{kT} n(x) \left(\frac{dE_{Fn}}{dx} - \frac{dE_i}{dx} \right) \quad (\text{A.11})$$

Using now (A.10), (A.11), and (1.3.18) in (A.9), we obtain

$$I_n(x) = A\mu_n n(x) \frac{dE_{Fn}}{dx} \quad (\text{A.12})$$

† E_{Fn} and E_{Fp} can be introduced better in the context of thermodynamics.

‡ If desired, one can let A shrink to zero around a point and define a *current density* at that point as dI/dA .

Thus the spatial variation of the electron quasi-Fermi level depends on the *total* electron current (drift plus diffusion components). Zero total electron current implies a constant electron quasi-Fermi level, and vice versa.

In a similar manner, the total hole current can be expressed in terms of drift and diffusion components by

$$I_p(x) = qA \left[-\mu_p p(x) \frac{d\psi}{dx} - D_p \frac{dp}{dx} \right] \quad (\text{A.13})$$

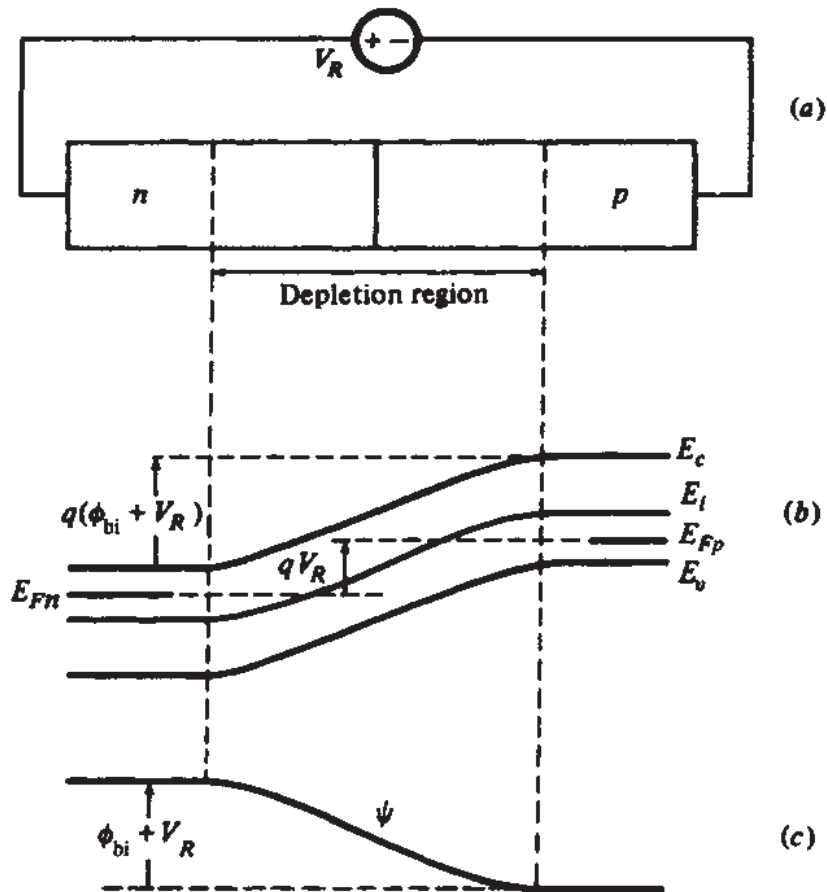
with μ_p being the hole mobility and D_p the hole diffusion constant given by $D_p = \mu_p \phi$, in analogy with (1.3.18). The algebraic signs in the above equation are easily discernible from the discussion in Secs. 1.3.2 and 1.3.3, remembering that the current direction considered there was opposite from the one here. Using (A.8) and proceeding as above, we obtain

$$I_p(x) = A\mu_p p(x) \frac{dE_{Fp}}{dx} \quad (\text{A.14})$$

which gives the relation between the *total* hole current (drift plus diffusion components) and the hole quasi-Fermi potential. Zero hole current implies a constant E_{Fp} , and vice versa.

Since in thermal equilibrium $E_{Fn} = E_{Fp} = E_F$ and E_F is constant (see Fig. A.4b), it is clear from (A.12) and (A.14) that thermal equilibrium implies zero total electron current *and* zero total hole current.

Quasi-Fermi level differences in semiconductor devices are often related to externally applied voltages. As an example, consider a *pn* junction with reverse bias V_R applied (Sec. 1.5). In the regions away from the depletion region, the majority carrier concentrations remain practically at their equilibrium values. However, the electrostatic potential across the depletion region must increase by V_R , corresponding to an increase in the energy band bending by qV_R , compared to Fig. A.4. Thus the energy band diagram with reverse bias becomes as shown in Fig. A.7. Separate quasi-Fermi levels are used so that, despite the extra band bending, (A.7) and (A.8) can give majority carrier concentrations at the *n* and *p* side, respectively, which have practically the same values as for Fig. A.4. The two quasi-Fermi levels are in this case said to “split” by the amount qV_R . As shown in Fig. A.7, the quasi-Fermi levels E_{Fn} and E_{Fp} are almost constant with position since I_n and I_p in (A.12) and (A.14) are very small under reverse bias, whereas *n* and *p*, correspondingly, are very large. At other places the shape of the quasi-Fermi levels can be deduced based on a number of assumptions concerning the carrier concentrations and the detailed mechanisms associated with the reverse-bias current. For the purposes of this book such considerations are not essential. It is only remarked here that assumptions commonly made lead to the conclusion that the quasi-Fermi levels continue practically horizontal over the depletion region (see, for example, Refs. 4 and 5 in Chap. 1).

**FIGURE A.7**

A pn junction with related energy band diagram and potential under reverse bias V_R .

BIBLIOGRAPHY

R. F. Pierret, *Semiconductor Device Fundamentals*, Addison-Wesley, Reading, Mass., 1996.

APPENDIX B

BASIC LAWS OF ELECTROSTATICS IN ONE DIMENSION

B.1 RELATIONS BETWEEN CHARGE DENSITY, ELECTRIC FIELD, AND POTENTIAL

The equations below are useful for one-dimensional analysis. (That is, it is assumed that all quantities involved vary only with the horizontal dimension x ; they are constant over a plane perpendicular to the x dimension for any given x .) The following symbols will be used:

$\rho(x)$ = charge density per unit volume at point x .

$\mathcal{E}(x)$ = electric field intensity at point x (in V/cm), with the field defined in the *positive* x direction.

$\psi(x)$ = electric potential at point x with respect to an arbitrary reference.

ϵ = permittivity of material; it is equal to $k\epsilon_0$, where k is the dielectric constant of the material and ϵ_0 the permittivity of free space (8.854×10^{-14} F/cm). The material is assumed to be characterized by a single value of ϵ everywhere, unless noted otherwise.

The above quantities are related as follows¹ (Fig. B.1):

$$\frac{d\mathcal{E}}{dx} = \frac{\rho(x)}{\epsilon} \quad (\text{B.1})$$

$$\frac{d\psi}{dx} = -\mathcal{E}(x) \quad (\text{B.2})$$

or, in integral form:

$$\mathcal{E}(x_2) - \mathcal{E}(x_1) = \frac{1}{\epsilon} \int_{x_1}^{x_2} \rho(x) dx \quad (\text{B.1a})$$

$$\psi(x_2) - \psi(x_1) = -\int_{x_1}^{x_2} \mathcal{E}(x) dx \quad (\text{B.2a})$$

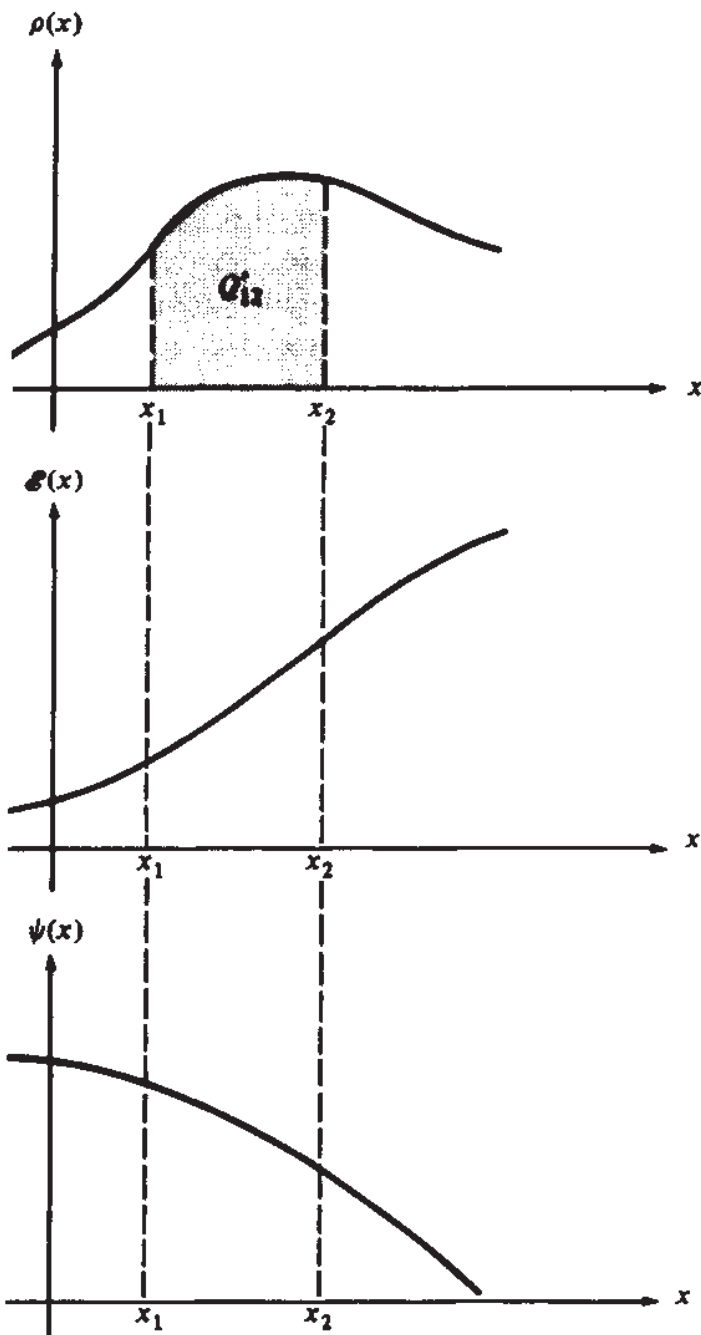


FIGURE B.1

Combining (B.1) with (B.2) we obtain Poisson's equation:

$$\frac{d^2\psi}{dx^2} = -\frac{\rho(x)}{\epsilon} \quad (\text{B.3})$$

B.2 RELATION BETWEEN ELECTRIC FIELD AND CHARGE PER UNIT AREA

Consider a parallelepiped as shown in Fig. B.2. The charge in a vanishingly thin, vertical slice of length Δx around point x will be $\rho(x)A \Delta x$. [As stated in the beginning of this appendix, $\rho(x)$ is assumed constant over any plane perpendicular to the x axis.] The total charge Q in the parallelepiped will be the integral of $\rho(x)A dx$ from x_1 to x_2 . Dividing this by A gives the charge per unit area Q' as seen from the side. Thus (B.1a) can be written

$$\mathcal{E}(x_2) - \mathcal{E}(x_1) = \frac{Q'_{12}}{\epsilon} \quad (\text{B.4})$$

where Q'_{12} is the charge per unit area contained between two vertical planes at x_1 and x_2 , as shown in Fig. B.1. The above equation is referred to as *Gauss' law*.

If points x_1 and x_2 above belong to two different materials joined at a plane perpendicular to the x dimension at a point x_0 somewhere between x_1 and x_2 , and if ϵ_1 and ϵ_2 are the corresponding permittivities, we have, in lieu of (B.4),

$$\epsilon_2 \mathcal{E}(x_2) - \epsilon_1 \mathcal{E}(x_1) = Q'_{12} \quad (\text{B.5})$$

B.3 DISCONTINUITIES IN ELECTRIC FIELD

As can be deduced from (B.5) by letting x_1 and x_2 approach the same value x_0 , the electric field can be discontinuous at $x = x_0$ because of (1) change of permittivity at

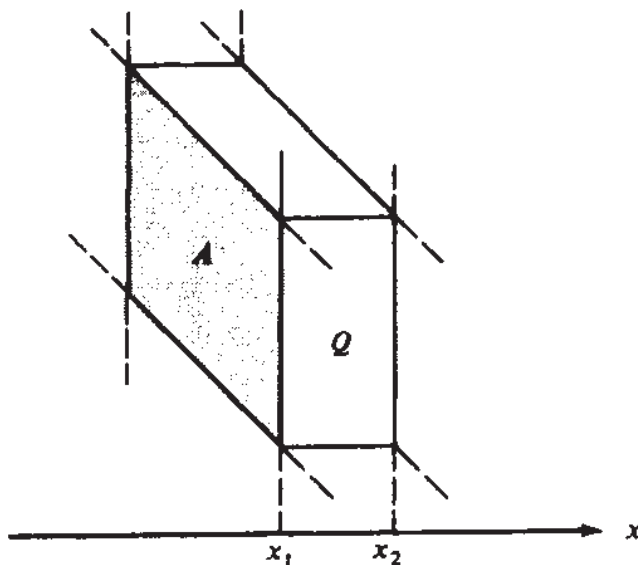


FIGURE B.2

an interface and/or (2) a charge sheet of zero thickness located at point x_0 , with a nonzero charge per unit area.

From the equations given above, one can derive the following two useful results that can be applied in several cases considered in this book.

B.4 RESULT 1

Consider a region characterized by a *uniform* charge density ρ_o and permittivity ϵ enclosed between two planes F and G perpendicular to the x dimension, as shown in Fig. B.3, separated by a distance d . Assume $\mathcal{E} = \mathcal{E}_F$ at the left plane. Then the potential drop ψ_{FG} between the two planes is given by

$$\psi_{FG} = d\mathcal{E}_F + \frac{\rho_o d^2}{2\epsilon} \quad (\text{B.6})$$

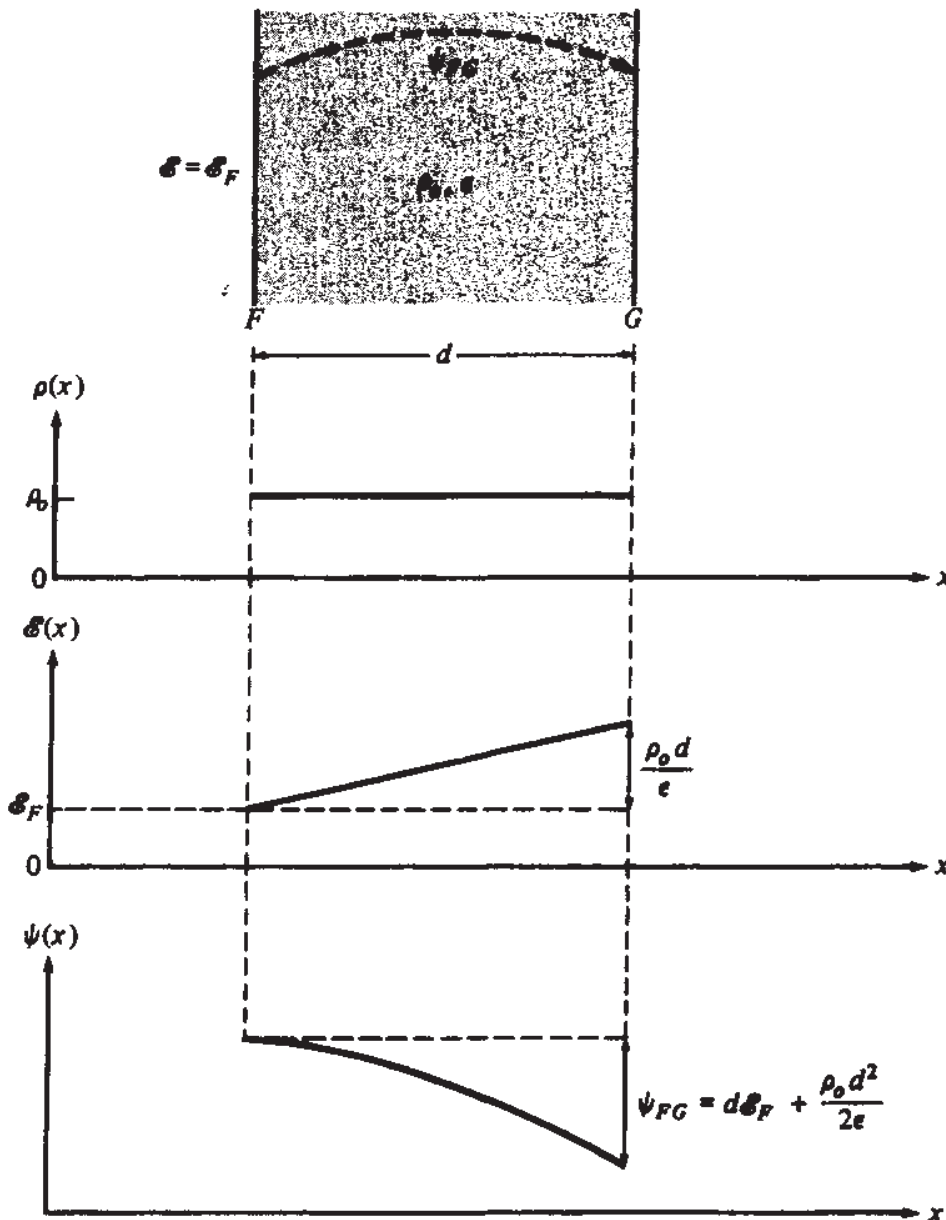


FIGURE B.3

The above result is a straightforward application of (B.1) and (B.2). The plots in Fig. B.3 are shown only for points between planes F and G .

B.5 RESULT 2

Consider the situation pictured in Fig. B.4. A region with permittivity ϵ and *no* charge in it (region II) is adjacent to a region with some charge in it (region I). The two regions are assumed separated by a plane perpendicular to the x dimension at point L as shown. Assume $\mathcal{E} = 0$ to the left of region I potential drop

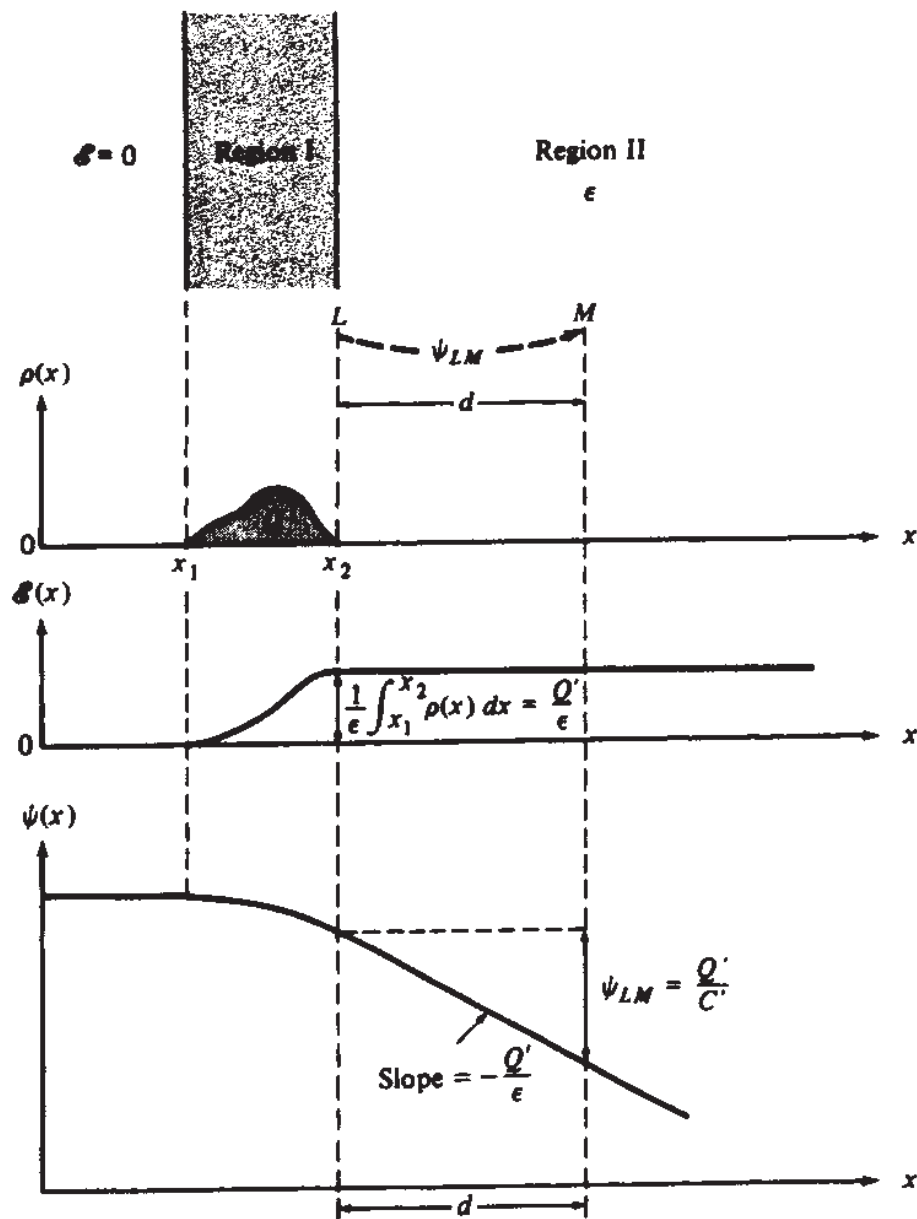


FIGURE B.4

from the boundary between the two regions to a point at a distance d inside region II will be

$$\psi_{LM} = \frac{Q'}{C'} \quad (\text{B.7})$$

where Q' is the charge per unit area in region I as seen from the side and C' is the capacitance per unit area between the planes at L and M , given by

$$C' = \frac{\epsilon}{d} \quad (\text{B.8})$$

The above result is again an application of (B.1) and (B.2). Note that the result in (B.7) is *independent* of the details of the shape of $\rho(x)$ in region I.

For an example of an application of the equations in this appendix, the reader is referred to Appendix C.

REFERENCE

1. J. D. Kraus, *Electromagnetics*, McGraw-Hill, New York, 1992.

APPENDIX C

CHARGE DENSITY, ELECTRIC FIELD, AND POTENTIAL IN THE *pn* JUNCTION

The equations in Appendix B can be employed in the analysis of a *pn* junction (Sec. 1.5), as shown in Fig. C.1. The external bias is assumed to be zero. The depletion approximation is used (Sec. 1.5). $\mathcal{E} = 0$ is assumed to the left of the depletion region. The corresponding energy band diagram is shown in Fig. A.4*b*. Note in Fig. C.1 that the contact potential of the *n* side to the *p* side is positive, as expected from the discussion in Sec. 1.5. From the relations shown on the plots one easily finds (Prob. 1.13)

$$\frac{l_1}{l_2} = \frac{N_A}{N_D} \quad (\text{C.1})$$

$$l_1 + l_2 = \sqrt{\frac{2\epsilon_s}{q} \frac{N_A + N_D}{N_A N_D} \phi_{bi}} \quad (\text{C.2})$$

The case handled here is that of a “two-sided” abrupt junction, in which N_D and N_A have comparable values. For the “one-sided” case, in which $N_D \gg N_A$, we have $l_1 \ll l_2$, and practically all of the potential drop ϕ_{bi} occurs across the *p* side. If the external bias is V_R , the same analysis is valid if ϕ_{bi} is replaced by $\phi_{bi} + V_R$.

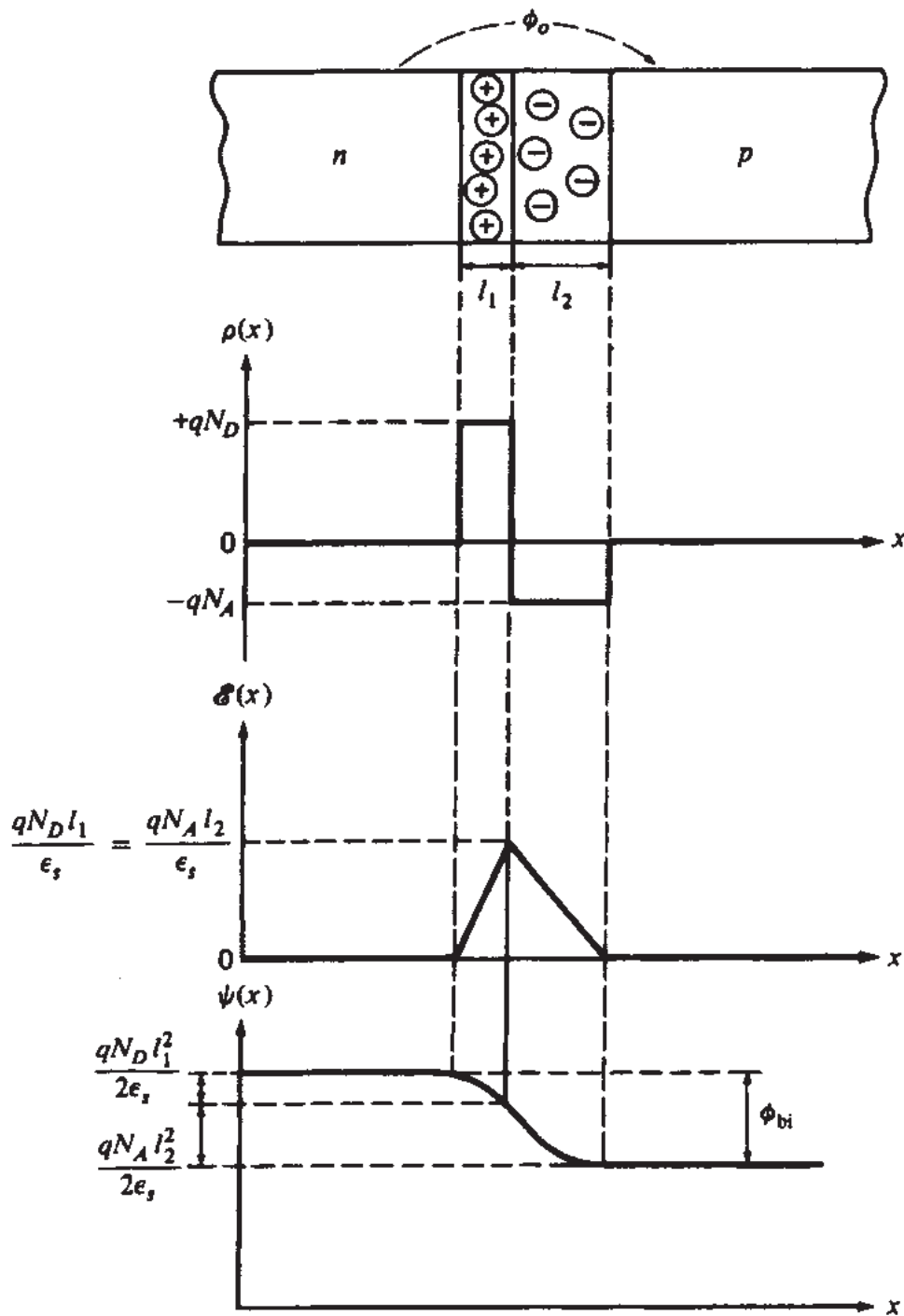


FIGURE C.1

BIBLIOGRAPHY

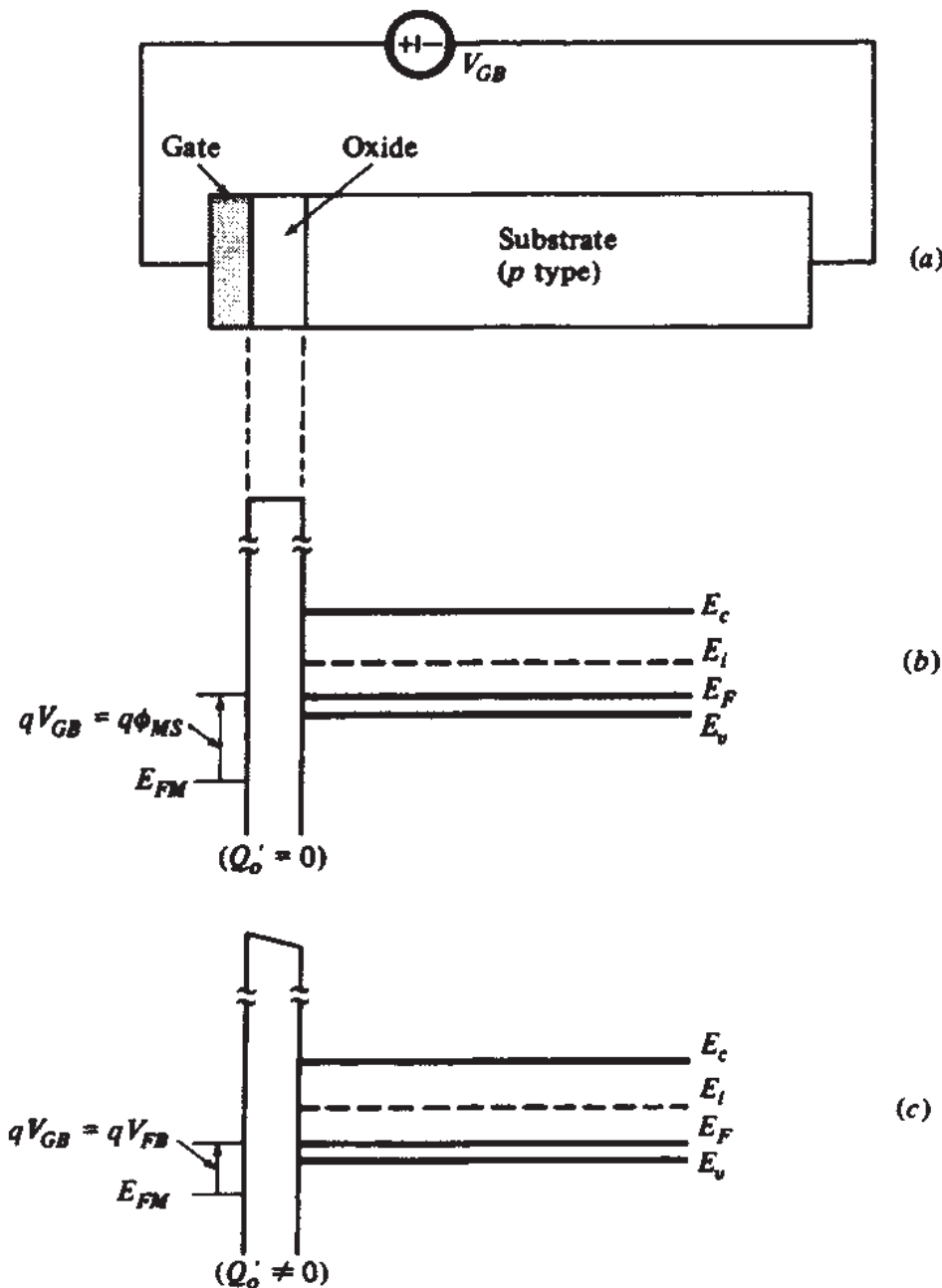
G. W. Neudeck, *The PN Junction Diode*, Addison-Wesley, Reading Mass., 1989.

APPENDIX D

ENERGY BAND DIAGRAMS FOR THE TWO-TERMINAL MOS STRUCTURE

The energy band model is often used in the description of MOS structures (see, for example, Refs. 1 to 4 in Chap. 2). The main features of such a description are summarized below. This material can best be understood by referring to Chap. 2, Secs. 2.2 to 2.5, and Appendix A.

Consider a two-terminal MOS structure on a p -type substrate, as shown in Fig. D.1a. Assume for the present that the effective interface charge per unit area, Q'_o , is zero. Let W_M and W_S be the work functions of the gate and substrate materials, respectively (Appendix A). As an example, assume $W_M > W_S$. If the MOS structure is short-circuited ($V_{GB} = 0$), electrons in the substrate, where the work function is lower, will find it attractive to leave and, traveling through the external wire, enter the gate. This will leave behind positive charges in the substrate, as has been shown in Fig. 2.2b, and will cause in it a potential drop. The other energy levels will thus bend until the Fermi levels in the gate and substrate can line up (as in the case of the pn junction in Fig. A.4). If it is desired to prevent this from happening and, instead, keep the bands "flat" in the substrate region next to the oxide, no potential drop and no net charges should exist there. This can be achieved by inserting in the external circuit a battery of value $V_{GB} = (W_M - W_S)/q$. When the connection is first established, the electrons in the substrate looking toward the external circuit face no longer the gate with its electrons of lower energy directly, but rather the negative terminal of the battery. The battery has increased the electron energy at that point by $W_M - W_S$. Now the

**FIGURE D.1**

(a) A two-terminal MOS structure with a p -type substrate; (b) energy band diagram for the flat-band condition, assuming $Q'_o = 0$; (c) energy band diagram for the flat-band condition, assuming $Q'_o \neq 0$.

environment the electrons would face if they attempted to leave the substrate is no more attractive than their present environment (the substrate itself). Thus there is no reason for electrons to leave and for charges to pile up. The external voltage source keeps the Fermi levels of the gate (E_{FM}) and of the substrate (E_F) separated by $E_F - E_{FM} = W_M - W_S$, and achieves the so-called *flat-band* condition shown in Fig. D.1b, which corresponds to Fig. 2.2c. A conduction band can be defined for the oxide and is shown to be horizontal since there is no field in the oxide under our assumption of $Q'_o = 0$. If now $Q'_o \neq 0$, then to keep the semiconductor bands flat, one has to adjust the external voltage to the value $V_{GB} = V_{FB}$ as given by (2.2.6). Now the conduction

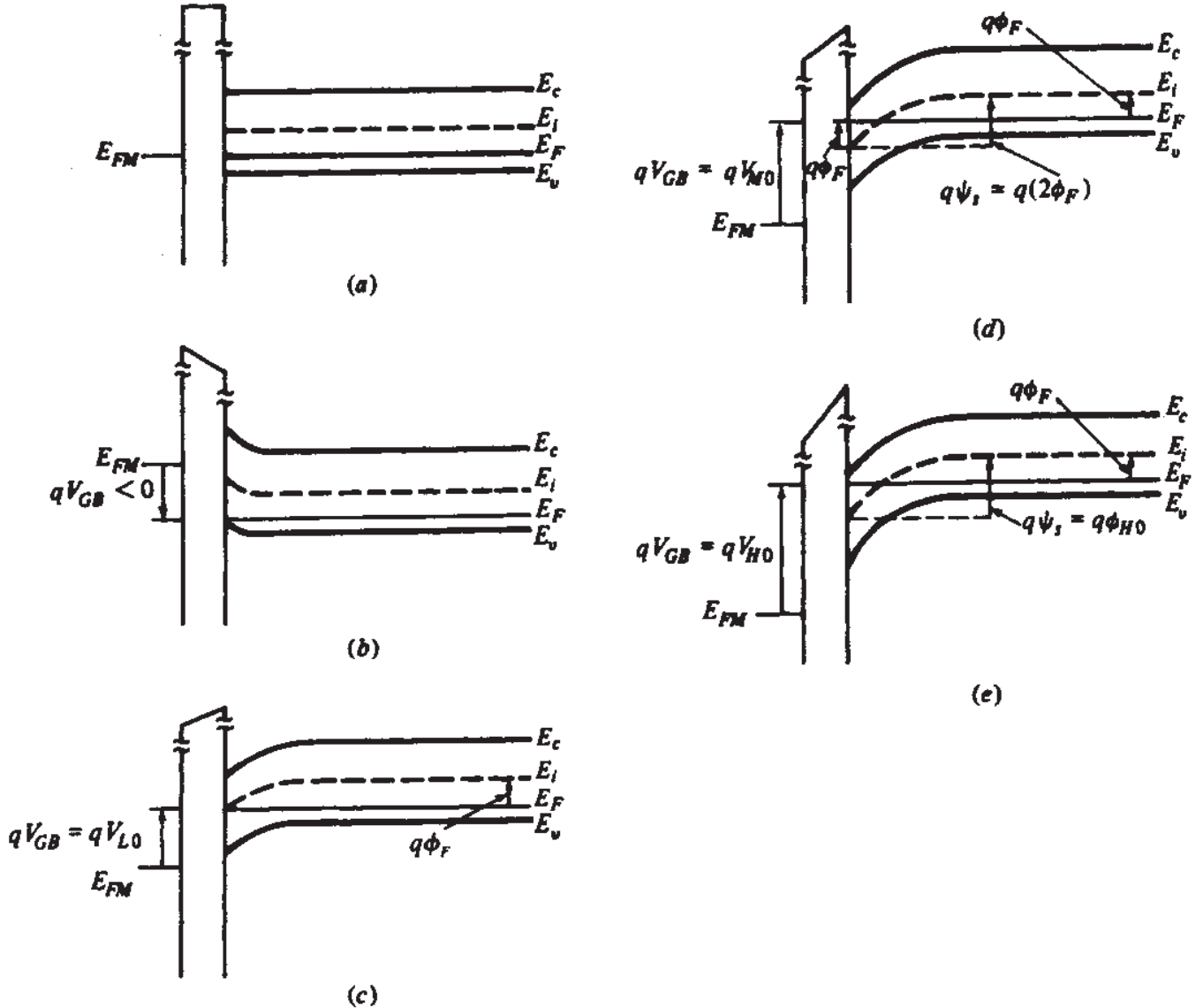
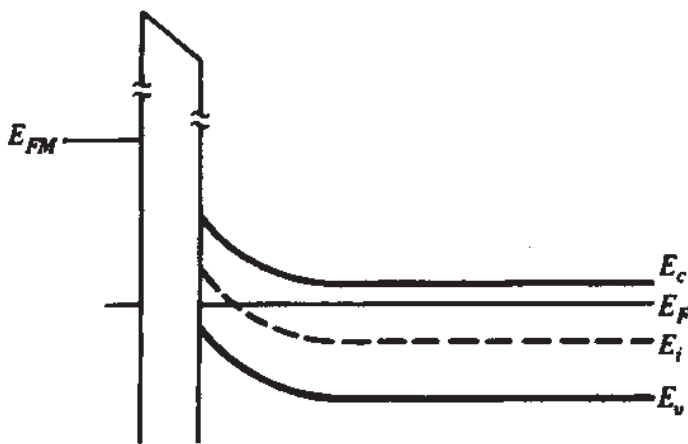


FIGURE D.2

Energy band diagrams for a two-terminal MOS structure with a p -type substrate, assuming $\phi_{MS} = 0$ and $Q'_o = 0$, for various values of V_{GB} (symbols are defined in Secs. 2.2 to 2.5). (a) Flat-band condition; (b) accumulation; (c) onset of weak inversion; (d) onset of moderate inversion; (e) onset of strong inversion.

band in the oxide will not be horizontal since there will be a nonzero oxide field (see Fig. 2.2e). The corresponding band diagram is shown in Fig. D.1c.

For simplicity in the rest of the discussion, we will assume both $\phi_{MS} = 0$ and $Q'_o = 0$. Then the flat-band condition corresponds to the band diagram shown in Fig. D.2a. Since now $V_{FB} = 0$, accumulation will be caused by $V_{GB} < 0$, and depletion or inversion by $V_{GB} > 0$. The band diagrams for various critical values of V_{GB} (defined in Secs. 2.4 and 2.5) are shown in Fig. D.2b to e. As expected from Appendix A, the surface potential ψ_s can be related to the bending of any level among E_c , E_i , E_v , as shown. The diagrams are self-explanatory.

**FIGURE D.3**

Energy band diagram for a two-terminal MOS structure with an n -type substrate at the onset of moderate inversion.

The construction of band diagrams for MOS structures with n -type substrates follows along the same lines. As an example, at the onset of moderate inversion we will have the case illustrated in Fig. D.3.

BIBLIOGRAPHY

- E. H. Nicollian and J. R. Brews, *MOS Physics and Technology*, John Wiley, New York, 1982.
S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.

APPENDIX E

CHARGE DENSITY, ELECTRIC FIELD, AND POTENTIAL IN THE TWO-TERMINAL MOS STRUCTURE

Figure E.1 shows the results of applying the basic laws of electrostatics given in Appendix B to the two-terminal MOS structure discussed in Chap. 2. The figure is based on three simplifying assumptions:

1. The depletion region has a uniform charge density and is assumed to contain only ionized acceptor atoms, and the region's bottom edge is sharply defined, with the semiconductor being neutral below it (depletion approximation).
2. The charges on the gate and at the interface are contained in extremely thin layers. The integral under the corresponding portions of the charge density plot (i.e., the corresponding charge per unit area) is Q'_G and Q'_o , respectively, as shown.
3. The inversion layer is much thinner than the depletion region. For simplicity, we assume that the charge density $\rho(y)$ is constant in the inversion layer. The latter assumption is not valid in practice. However, it can be seen from the material in Appendix B that, as the thickness of the inversion layer is allowed to approach zero (charge sheet model), the details of the shape of $\rho(y)$ in it become irrelevant anyway. All that counts, then, is the corresponding integral of the inversion layer charge density, which in Fig. E.1 has been denoted by Q'_I .

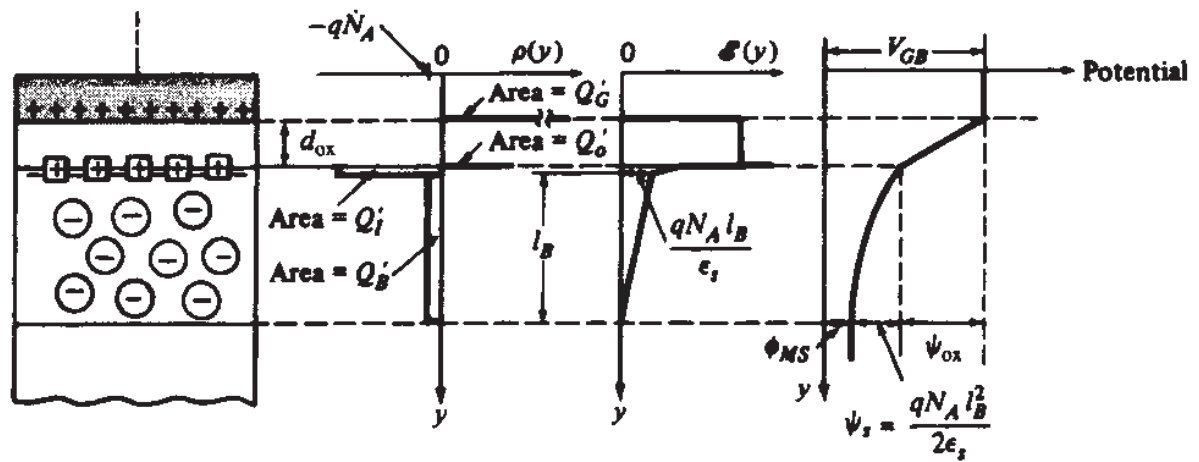


FIGURE E.1

BIBLIOGRAPHY

- E. H. Nicollian and J. R. Brews, *MOS Physics and Technology*, John Wiley, New York, 1982.
 S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.

APPENDIX F

GENERAL ANALYSIS OF THE TWO-TERMINAL MOS STRUCTURE

In our discussion of the two-terminal MOS structure in Sec. 2.5, we have focused on inversion and have adopted the charge sheet and depletion approximations. As is mentioned in Sec. 2.4, though, it is possible to analyze the structure without making such assumptions, allowing for the presence of both electrons and holes throughout the semiconductor, with distributions dictated from semiconductor physics. Such analysis will be valid in all operating regimes (accumulation, depletion, and inversion). In inversion, it accounts for the fact that the hole concentration decreases continuously as one goes from the bulk toward the surface, and allows for the spreading of the inversion layer below the surface. Here we provide the main steps and results of this general analysis. In using relations from Appendix B, x will be replaced by y since quantities vary in the vertical direction, assuming the orientation we have adopted in drawing the MOS two-terminal structure in Chap. 2. A p -type substrate is assumed.

Let us refer to Fig. 2.4a. The charge density at depth y in the semiconductor, including holes, electrons, and ionized acceptor atoms, is given by (2.4.15) repeated here for convenience:

$$\rho(y) = q[p(y) - n(y) - N_A] \quad (\text{F.1})$$

From (2.4.13) and (2.4.14) we have

$$n(y) = n_o e^{\psi(y)/\phi_t} \quad (\text{F.2})$$

$$p(y) = p_o e^{-\psi(y)/\phi_t} \quad (\text{F.3})$$

where $\psi(y)$ is the potential with respect to the bulk at y . Deep in the bulk the charge density is zero, so (F.1) gives $p_o - n_o = N_A$. Using N_A from this and also (F.2) and (F.3) in (F.1), and the result in Poisson's equation (Appendix B), we have

$$\frac{d^2\psi}{dy^2} = -\frac{q}{\epsilon_s} \left[p_o \left(e^{-\psi(y)/\phi_t} - 1 \right) - n_o \left(e^{\psi(y)/\phi_t} - 1 \right) \right] \quad (\text{F.4})$$

Assuming $N_A \gg n_i$, we can use (1.2.4) and (1.2.5) in the above equation. Eliminating n_i in the result by using (1.4.2), we obtain

$$\frac{d^2\psi}{dy^2} = -\frac{qN_A}{\epsilon_s} \left[e^{-\psi(y)/\phi_t} - 1 - e^{-2\phi_F/\phi_t} \left(e^{\psi(y)/\phi_t} - 1 \right) \right] \quad (\text{F.5})$$

Multiply both sides of this equation by $2(d\psi/dy)$; the resulting left-hand side can be recognized as $(d/dy)(d\psi/dy)^2$. Replace y by a dummy variable \hat{y} and integrate from a point deep in the bulk (theoretically at infinity, where $\psi = 0$ and $d\psi/dy = 0$) to a point y . Solve for $d\psi/dy$ at point y , and recall that $\mathcal{E}(y) = -d\psi/dy$ (Appendix B). The following is then obtained:

$$\mathcal{E}(y) = -\frac{d\psi}{dy} = \pm \frac{\sqrt{2q\epsilon_s N_A}}{\epsilon_s} \sqrt{\phi_t e^{-\psi/\phi_t} + \psi - \phi_t + e^{-2\phi_F/\phi_t} (\phi_t e^{\psi/\phi_t} - \psi - \phi_t)} \quad (\text{F.6})$$

where $\psi = \psi(y)$ and the + sign in front of the right-hand side is to be used with $\psi > 0$, and the - sign with $\psi < 0$. This choice of signs corresponds to the fact that the signs of \mathcal{E} and ψ agree, as follows from the discussion in Sec. 2.4. To find the total semiconductor charge per unit area Q'_C , we can apply (B.4) of Appendix B, taking point 1 at the surface and point 2 deep in the bulk, where $\mathcal{E} = 0$. This gives $-\mathcal{E}_{\text{surface}} = Q'_C/\epsilon_s$. Evaluating $\mathcal{E}_{\text{surface}}$ from (F.6) with $\psi = \psi_s$ (the surface potential), we obtain:

$$Q'_C = \mp \sqrt{2q\epsilon_s N_A} \sqrt{\phi_t e^{-\psi_s/\phi_t} + \psi_s - \phi_t + e^{-2\phi_F/\phi_t} (\phi_t e^{\psi_s/\phi_t} - \psi_s - \phi_t)} \quad (\text{F.7})$$

This function behaves as shown in Fig. F.1.

Using the above equation in (2.6.7) gives

$$C'_c = \pm \sqrt{2q\epsilon_s N_A} \left\{ \frac{1 - e^{-\psi_s/\phi_t} + e^{-2\phi_F/\phi_t} (e^{\psi_s/\phi_t} - 1)}{2 \sqrt{\phi_t e^{-\psi_s/\phi_t} + \psi_s - \phi_t + e^{-2\phi_F/\phi_t} (\phi_t e^{\psi_s/\phi_t} - \psi_s - \phi_t)}} \right\} \quad (\text{F.8})$$

Equations (F.7) and (F.8) are valid in all regions (accumulation, depletion, and inversion). It is easy to see that in each region certain terms can be neglected. When using (F.8) with very small $|\psi_s|$, one should watch out for numerical inaccuracies because both its numerator and denominator go to zero as ψ_s goes to zero.

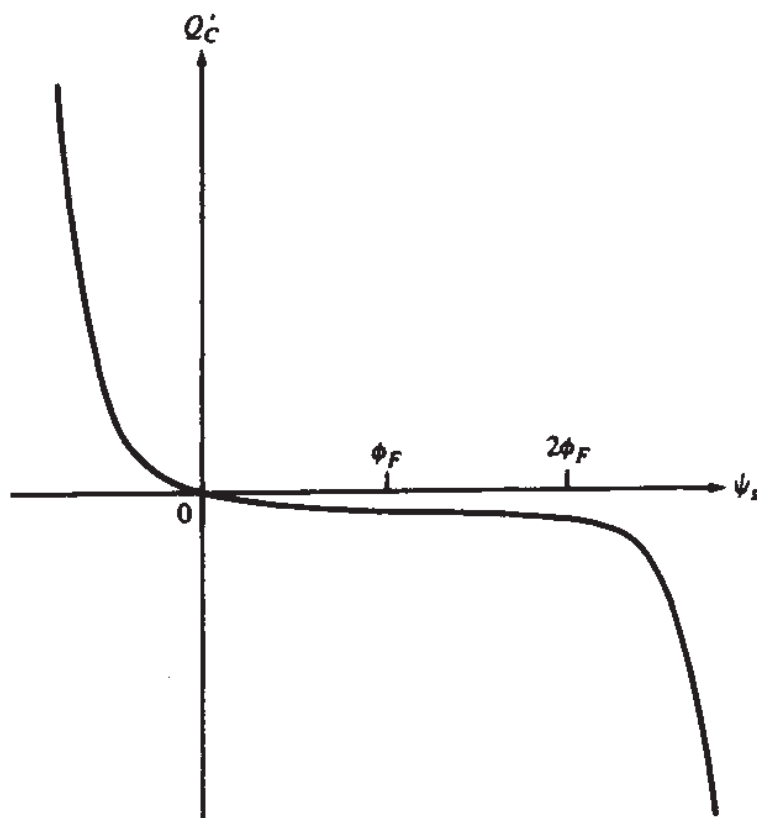


FIGURE F.1

The relation between ψ and y can be obtained from (F.6) by separating variables and integrating from a point at y to a point at the surface. This gives

$$\int_{\psi(y)}^{\psi_s} \frac{d\hat{\psi}}{\mathcal{E}[\hat{\psi}]} = y - y_{\text{surface}} \quad (\text{F.9})$$

where $\hat{\psi}$ is a dummy variable of integration, and $\mathcal{E}[\psi]$ is given by (F.6). From this one can determine ψ for a given y (and a given ψ_s) numerically. With $\psi(y)$ obtained in this way, one can then determine $\rho(y)$, $n(y)$, $p(y)$, and $\mathcal{E}(y)$ from (F.1) to (F.3) and (F.6).

The rest of the analysis proceeds as explained in Sec. 2.4.4.

INVERSION

In inversion it is easy to see that (F.7) reduces practically to (2.5.1), and (F.8) reduces to (2.6.9). The charges and capacitances (per unit area) corresponding to the inversion region and the depletion region can be determined as follows. In (2.5.4) we use $n(y)$ from (F.2) and in the result we use (1.4.1). We take y_c deep in the bulk (theoretically at infinity, where $\psi = 0$). We perform a change of variables from y to ψ . The re-

sulting factor $(d\psi/dy)^{-1}$ in the integrand can be obtained from (F.6). This procedure gives:

$$Q'_I = -qN_A e^{-2\phi_F/\phi_t} \int_0^{\psi_s} \frac{e^{\psi/\phi_t}}{\mathcal{E}[\psi]} d\psi \quad (\text{F.10})$$

A similar procedure gives the bulk charge per unit area Q'_B , consisting of ionized acceptor atoms and holes:

$$Q'_B = -qN_A \int_0^{\psi_s} \frac{1 - e^{-\psi/\phi_t}}{\mathcal{E}[\psi]} d\psi \quad (\text{F.11})$$

The above integrals can be evaluated numerically. However, determining the corresponding capacitances is easy. Using the above expressions in (2.6.12) and (2.6.13) and performing the differentiation results in *explicit* expressions for C'_I and C'_B . After dropping the terms that are negligible in the inversion region, we easily obtain (2.6.14) and (2.6.15).

Sometimes an intermediate level of approximation is used: the charge sheet assumption (i.e., infinitesimal thickness) for the inversion layer is taken to hold, but holes are *still* allowed to exist in the depletion region. Then $n(y)$ is taken as 0 in (F.1) at any point below the inversion layer. Thus the second term in the brackets in (F.4) will be zero as will the fourth term in the sum in (F.6). The *bulk* charge per unit area Q'_B can then be found by applying (B.4) as above, taking point 2 again in the bulk but point 1 immediately below the inversion layer. This gives

$$Q'_B \approx -\sqrt{2q\epsilon_s N_A} \sqrt{\psi_s - \phi_t} \quad (\text{F.12})$$

Finally, if the depletion approximation is used and thus not even holes are allowed in the depletion region, the only term that will be present in the right-hand side of (F.1) (for any y below the inversion layer) is N_A . Using this directly in Poisson's equation and integrating gives (2.5.6), which is the approximation widely used in this book.

BIBLIOGRAPHY

- E. H. Nicollian and J. R. Brews, *MOS Physics and Technology*, John Wiley, New York, 1982.
S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.

APPENDIX G

CAREFUL DEFINITIONS FOR THE LIMITS OF MODERATE INVERSION

TWO-TERMINAL STRUCTURE. Throughout this appendix, uniform substrates are assumed. In Sec. 2.5, the lower limit of the moderate-inversion region was taken to coincide with the upper limit of weak inversion. The latter is traditionally defined at $\psi_s = 2\phi_F$. At this point, the surface electron concentration becomes equal to the bulk doping concentration N_A , as seen from (2.4.11*b*). Also, at this point $C'_i = C'_b$, as seen from (2.6.14) and (2.6.15). These two facts hold independently of the value of C'_{ox} . However, neither of these facts says anything about the validity of common “weak-inversion approximations” [such as $Q'_i(V_{GB})$ being an exponential] at this point. This is because the validity of such approximations cannot be discussed carefully without evoking the value of C'_{ox} . For detailed work, then, it makes practical sense to take C'_{ox} into account, and redefine the onset of moderate inversion at some point beyond which common weak inversion approximations become unacceptable. In weak inversion, and at points where C'_i is negligible (Fig. 2.16), C'_b varies little; thus, $d\psi_s/dV_{GB}$ in (2.6.18) will be approximately constant. That is necessary for an exponential dependence of Q'_i on V_{GB} , as can be seen from the development leading from (2.5.36) to (2.5.42). Significant departure from such behavior will be observed if C'_i starts becoming significant in comparison to $C'_{ox} + C'_b$ in (2.6.18). Thus, let us redefine the onset of moderate inversion as follows:¹

ϕ_{M0} and \hat{V}_{M0} are the values of ψ_s and V_{GB} , respectively, at which

$$\frac{C'_i}{C'_{ox} + C'_b} = 0.1 \quad (G.1)$$

Keeping in mind that C'_b varies slowly, and using (2.6.18), it is easy to see that at this point the slope $d\psi_s/dV_{GB}$ in Fig. 2.9 drops to about 91 percent of its value deep in weak inversion. The same is true for the slope of $\ln |Q'_I|$ vs. V_{GB} (Fig. 2.12), this slope being a measure of the "exponentiality" of $Q'_I(V_{GB})$.

Accurate evaluation of ϕ_{M0} requires using (2.6.14) and (2.6.15) in (G.1), and solving iteratively for ψ_s ; the corresponding \hat{V}_{M0} can then be found from (2.5.15). The result depends on oxide thickness and substrate doping. It can be shown that, for practical values of these parameters, the value of ϕ_{M0} as defined above will not differ by more than about $1\phi_t$ from $2\phi_F$. The corresponding value of \hat{V}_{M0} will differ from that given in (2.5.18) by at most a few tens of millivolts. For most practical cases, we can continue using the value of $2\phi_F$ and (2.5.18) for simplicity.

We now turn to the upper limit of moderate inversion. For detailed work, it makes practical sense to define this at a point below which common strong-inversion approximations, such as (2.5.26), become unacceptable. Let us look at the slope of the $Q'_I(V_{GB})$ plot, given by (2.6.20). Deep in strong inversion, C'_i is very large and (2.6.20) reduces to $d|Q'_I|/dV_{GB} \approx C'_{ox}$; this agrees of course with (2.5.26). If C'_i is not much larger than $C'_{ox} + C'_b$, $d|Q'_I|/dV_{GB}$ will be less than C'_{ox} and (2.5.26) will not hold. Thus, let us define the upper limit of moderate inversion as follows:¹

ϕ_{H0} and V_{H0} are the values of ψ_s and V_{GB} , respectively, at which

$$\frac{C'_i}{C'_{ox} + C'_b} = 10 \quad (G.2)$$

From (2.6.20) it is easy to see that at this point the slope in the plot of $|Q'_I|$ versus V_{GB} (Fig. 2.10) is reduced to about 91 percent of its theoretical maximum value of C'_{ox} and, from (2.6.18) it can be seen that the slope of the $\psi_s(V_{GB})$ plot (Fig. 2.9) drops to about 9 percent of its maximum value.

To find ϕ_{H0} one must use (2.6.14) and (2.6.15) in (G.2) and solve for ψ_s iteratively. One finds that ϕ_{H0} is several ϕ_t above ϕ_{M0} (about $6\phi_t$), the exact value being dependent on oxide thickness and substrate doping. If ϕ_{H0} is accurately known, it can be used in (2.5.15) to find V_{H0} . (Note that a small error in ϕ_{H0} will result in a large error in V_{H0} because of the exponential term in that equation; that term is now large.)

Using the above calculations, the width of the moderate-inversion region can be found to vary significantly with process parameters. However, not all combinations of oxide thickness and substrate doping are practical. For example, in MOS transistor fabrication large doping concentrations are usually combined with thin oxides. For practical cases, then, the width of the moderate inversion does not differ much; an average value is about 0.6 V at room temperature.

THREE-TERMINAL STRUCTURE. For the three-terminal MOS structure discussed in Chap. 3, the same rationale can be used. Thus, the surface potential ϕ_M at the boundary between weak and moderate inversion for a given V_{CB} can be found from (G.1), using in it (3.2.12) and (3.2.13) and solving iteratively for $\psi_s = \phi_M$. The solution turns out to be within about $1\phi_i$ of $2\phi_F$, depending on process parameters. The surface potential ϕ_H at the boundary between moderate and strong inversion can be found from (G.2) in a similar manner, and turns out to be $5\phi_i$ to $6\phi_i$ higher. The corresponding V_{GB} values can be found by using the above surface potential values in (3.2.7b). Calculated this way, the onset of moderate inversion turns out to be very close to the value of V_{MB} given in Table 3.1 (within a few tens of millivolts). The onset of strong inversion is above this by several tenths of 1 V. The width of moderate inversion in terms of V_{GB} tends to decrease somewhat with V_{CB} (e.g., by about 0.1 V or less, for V_{CB} increasing from 0 to 3 V, for practical devices).

Rather than finding ϕ_M and ϕ_H for a given V_{CB} , as suggested above, one can find these limits for a given V_{GB} . One can use (3.2.12) and (3.2.13) in the definitions (G.1) and (G.2), and then utilize (3.2.7b) to eliminate the exponential terms. The resulting equations can be solved *explicitly* for ϕ_M and ϕ_H . These values can then be substituted back in (3.2.7b) to find explicitly the corresponding V_{CB} values, V_W and V_Q respectively, for the given V_{GB} (Prob. 3.10).

REFERENCE

1. Y. Tsividis, "Moderate inversion in MOS devices," *Solid-State Electronics*, vol. 25, pp. 1099–1104, 1982; see also Erratum, *ibid.*, vol. 26, p. 823, 1983.

APPENDIX H

ENERGY BAND DIAGRAMS FOR THE THREE-TERMINAL MOS STRUCTURE

Consider the three-terminal MOS structure on a p -type substrate, discussed in Chap. 3 (Fig. H.1a). Assume first that $V_{CB} = 0$. As mentioned in Sec. 3.2, this leads to the same conditions as in a two-terminal MOS structure. For a certain surface potential ψ_s of value ψ_1 the energy band diagram in the semiconductor is as shown in Fig. H.1b (Appendix D). If now V_{GB} is kept unchanged and V_{CB} is made positive, the electron energy in the n^+ region will be lowered by qV_{CB} and the electrons will find this region more attractive. Their concentration in the inversion layer will be reduced. If it is desired to restore the surface to its former level of inversion, the energy at the surface must be lowered by qV_{CB} also. Thus the band bending at the surface must increase (in comparison to Fig. H.1b) by qV_{CB} (by increasing V_{GB}), resulting in the band diagram of Fig. H.1c. Now the electron concentration n at the surface has the original value. Since we now have nonequilibrium, one must use (A.7) to predict n . For this to give the same value as (A.1) gives for Fig. H.1b, the electron quasi-Fermi level E_{Fn} at the surface must maintain the same position relative to E_i as E_F does in Fig. H.1b. This is shown in Fig. H.1c. Deep in the bulk the hole quasi-Fermi level E_{Fp} is maintained at the same position relative to E_i as E_F is in Fig. H.1b, so that (A.8) will produce the same value for p as (A.2) did in equilibrium. From the above discussion it is seen that the E_{Fn} at the surface and E_{Fp} in the bulk split by qV_{CB} .

If no charge sheet approximation is made and the electron concentration $n(y)$ is allowed to be nonzero at point y below the surface (as in the general analysis outlined

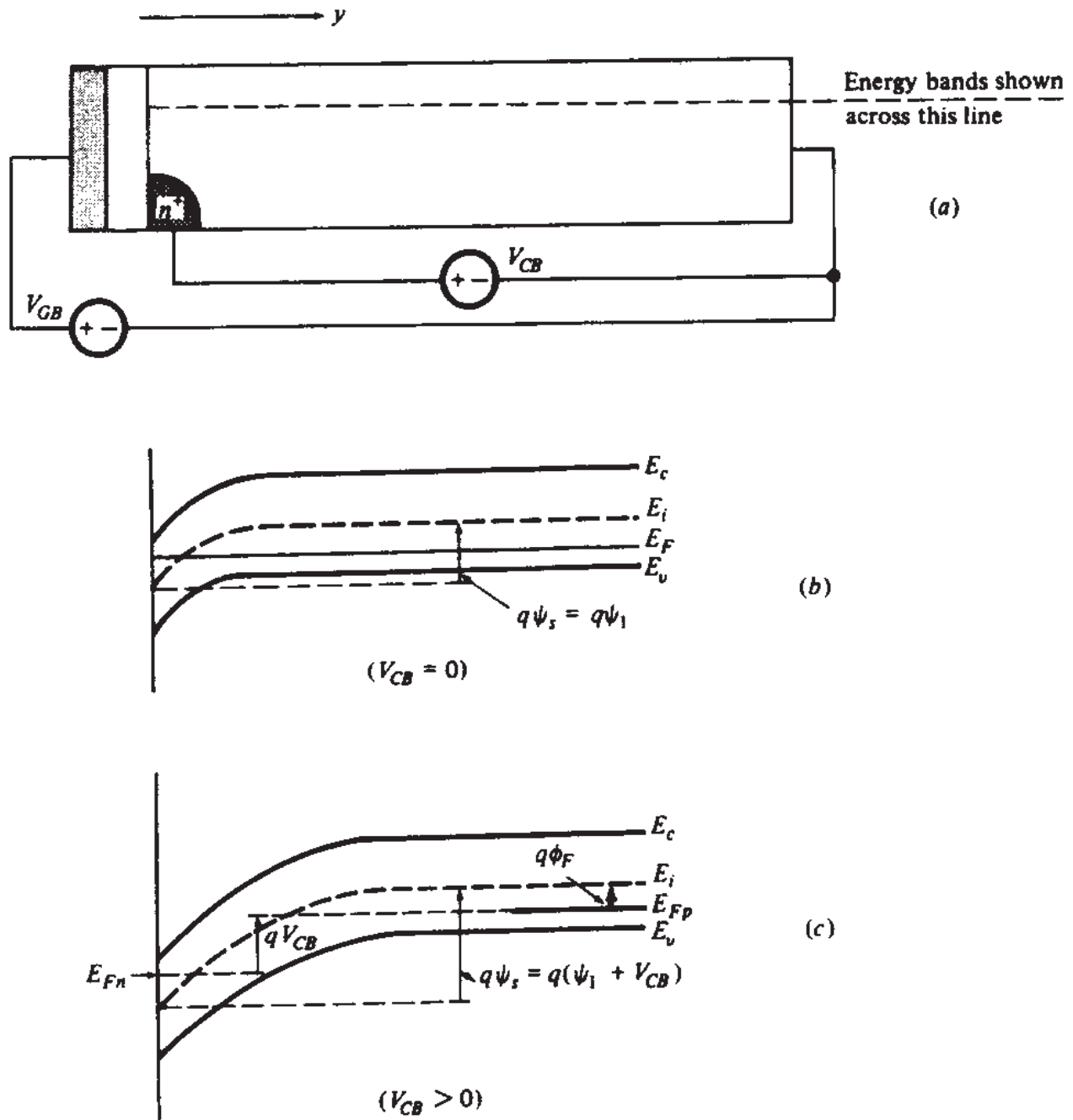


FIGURE H.1

in Sec. 2.4), one can extend the above arguments for point y . Thus, E_{Fn} is taken practically constant with y below the surface also.^{1,2} This, as follows from the material on quasi-Fermi levels in Appendix A, is equivalent to assuming that there is negligible electron current in the direction perpendicular to the surface. Indeed, as mentioned in Sec. 3.2, in our analysis we neglect the very small electron and hole currents responsible for the minute "reverse-bias" current flowing through the external source V_{CB} . To express $n(y)$ we can use (A.7) with $E_{Fn} - E_i(y)$ determined as shown in Fig. H.2; this gives

$$n(y) = n_i e^{[\psi(y) - \phi_F - V_{CB}]/\phi_1} \quad (\text{H.1})$$

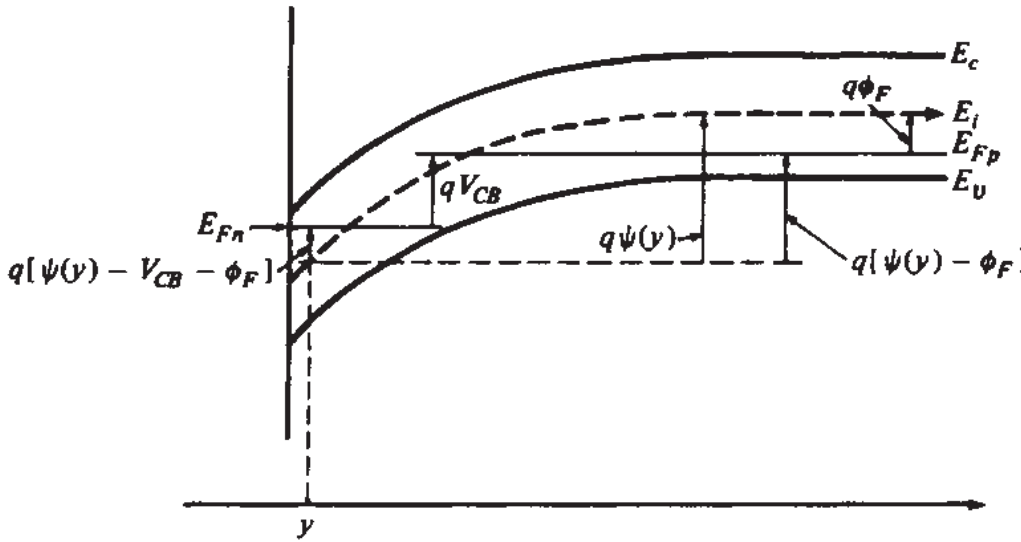


FIGURE H.2

with $\phi_t = kT/q$. Using (1.4.1a) and (1.4.1b), this can be written as follows:

$$n(y) = n_o e^{[\psi(y) - V_{CB}]/\phi_t} \quad (\text{H.2a})$$

$$= p_o e^{[\psi(y) - 2\phi_F - V_{CB}]/\phi_t} \quad (\text{H.2b})$$

and, from (1.2.4),

$$n(y) \approx N_A e^{[\psi(y) - 2\phi_F - V_{CB}]/\phi_t} \quad (\text{H.3})$$

At the surface, where $\psi(y) = \psi_s$, this equation reduces to (3.2.1).

To claim that the electron concentration deep in the bulk in Fig. H.1c is as that shown in Fig. H.1b, one would have to allow E_{Fn} to rise eventually as one moves toward the bulk, and assume deep in the bulk the same position relative to E_i as for Fig. H.1b (from Appendix A this would imply some flow of electron current). However, at such points away from the surface, the electron concentration is extremely small, and its exact value does not make any significant difference in the analysis. For simplicity, then, the above equations for $n(y)$ are left unmodified.

The quasi-Fermi level for holes, E_{Fp} , is commonly assumed to remain constant throughout the semiconductor, because the holes do not communicate with the external source V_{CB} in the sense discussed above for electrons. A constant E_{Fp} is consistent with our assumption of zero hole current. Using (A.8) for Fig. H.2, we obtain

$$p(y) = n_i e^{(\phi_F - \psi(y))/\phi_t} \quad (\text{H.4})$$

which, from (1.4.1b), becomes

$$p(y) = p_o e^{\psi(y)/\phi_t} \quad (\text{H.5})$$

Note that, even if E_{Fp} is not exactly constant so that $p(y)$ is not given exactly by the above equations, in inversion holes play a negligible role anyway; hence it is not worth complicating the above equations.

REFERENCES

1. H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator) semiconductor transistors," *Solid-State Electronics*, vol. 9, pp. 927-937, 1966.
2. S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.

APPENDIX I

GENERAL ANALYSIS OF THE THREE-TERMINAL MOS STRUCTURE

For the two-terminal MOS structure on a p -type substrate the electron concentration at y was given by (2.4.13). According to the discussion in Sec. 3.2, this equation can be converted to one valid for the p -substrate three-terminal MOS structure of Fig. 3.1c by replacing $\psi(y)$ by $\psi(y) - V_{CB}$:

$$n(y) = n_o e^{[\psi(y) - V_{CB}]/\phi_t} \quad (\text{I.1})$$

Assuming that the holes do not communicate directly with the external source in the sense discussed in Sec. 3.2 for electrons,^{1,2†} their concentration will depend only on $\psi(y)$, as was the case in (2.4.14) for the two-terminal structure:

$$p(y) = p_o e^{-\psi(y)/\phi_t} \quad (\text{I.2})$$

The above two equations have also been derived and discussed by using energy band concepts in Appendix H. Proceeding now as in Appendix F, we obtain Poisson's equation as follows:

$$\frac{d^2\psi}{dy^2} = -\frac{q}{\epsilon_s} \left[p_o \left(e^{-\psi(y)/\phi_t} - 1 \right) - n_o \left(e^{[\psi(y) - V_{CB}]/\phi_t} - 1 \right) \right] \quad (\text{I.3})$$

†This assumption is not exactly valid because both holes and electrons are responsible for the minute reverse-bias current mentioned in Sec. 3.2.

Following the procedure outlined in Appendix F, the solution of the above equation leads to the following results:

$$\begin{aligned}\mathcal{E}(y) &= \frac{d\psi}{dy} \\ &= \pm \frac{\sqrt{2q\epsilon_s N_A}}{\epsilon_s} \sqrt{\phi_i e^{-\psi/\phi_i} + \psi - \phi_i + e^{-2\phi_F/\phi_i} \left(\phi_i e^{(\psi-V_{CB})/\phi_i} - \psi - \phi_i e^{-V_{CB}/\phi_i} \right)}\end{aligned}\quad (I.4)$$

$$Q'_C = \mp \sqrt{2q\epsilon_s N_A} \sqrt{\phi_i e^{-\psi_s/\phi_i} + \psi_s - \phi_i + e^{-2\phi_F/\phi_i} \left(\phi_i e^{(\psi_s-V_{CB})/\phi_i} - \psi_s - \phi_i e^{-V_{CB}/\phi_i} \right)} \quad (I.5)$$

$$C'_c = \pm \sqrt{2q\epsilon_s N_A} \frac{1 - e^{-\psi_s/\phi_i} + e^{-2\phi_F/\phi_i} \left(e^{(\psi_s-V_{CB})/\phi_i} - 1 \right)}{2 \sqrt{\phi_i e^{-\psi_s/\phi_i} + \psi_s - \phi_i + e^{-2\phi_F/\phi_i} \left(\phi_i e^{(\psi_s-V_{CB})/\phi_i} - \psi_s - \phi_i e^{-V_{CB}/\phi_i} \right)}} \quad (I.6)$$

$$\int_{\psi(y)}^{\psi_s} \frac{d\hat{\psi}}{\mathcal{E}[\hat{\psi}]} = y - y_{\text{surface}} \quad (I.7)$$

where $\mathcal{E}[\psi]$ is given by the right-hand side of (I.4).

INVERSION

For the inversion region, proceeding as in Appendix F we obtain:

$$Q'_I = -qN_A e^{(-2\phi_F-V_{CB})/\phi_i} \int_0^{\psi_s} \frac{e^{\psi/\phi_i}}{\mathcal{E}[\psi]} d\psi \quad (I.8)$$

$$Q'_B = -qN_A \int_0^{\psi_s} \frac{1 - e^{-\psi/\phi_i}}{\mathcal{E}[\psi]} d\psi \quad (I.9)$$

Deleting negligible terms in (I.5) and using (2.5.2) and (3.2.5a) leads to (3.2.6). Similarly, deleting negligible terms in (I.6) leads to (3.2.11). Finally, using (I.8) in (2.6.13) and (I.9) in (2.6.12) results in explicit expressions for C'_i and C'_b . From these expressions, after deleting terms that are negligible in inversion, we obtain (3.2.12) and (3.2.13).

REFERENCES

1. H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator) semiconductor transistors," *Solid-State Electronics*, vol. 9, pp. 927-937, 1966.
2. S. M. Sze, *Physics of Semiconductor Devices*, John Wiley, New York, 1981.

APPENDIX J

DRAIN CURRENT FORMULATION USING QUASI-FERMI POTENTIALS

In Sec. 4.3 we evaluated the drain current caused by both drift and diffusion. The current components caused by each of these two phenomena were kept separate. In this appendix we present an alternative approach, resulting in a compact formulation which combines the two effects.¹ We consider the flow of electrons as laminar flow. That is, we assume that the total current I_{DS} is the sum of elemental currents ΔI , each flowing horizontally in an inversion layer slice parallel to the surface, having width W and depth Δy , centered at point y , as shown in Fig. J.1. In the general case, we will have for the current in each slice, allowing Δy to become a differential,

$$dI_{DS} = dI_{\text{drift}}(x, y) + dI_{\text{diff}}(x, y) \quad (\text{J.1})$$

where it should be kept in mind that our direction for the current is from right to left, as shown in Fig. J.1. The drift component is, from (1.3.11),

$$dI_{\text{drift}}(x, y) = (W dy) q \mu n(x, y) \frac{\partial \psi(x, y)}{\partial x} \quad (\text{J.2})$$

The diffusion component is, from (1.3.17) and (1.3.18),

$$dI_{\text{diff}}(x, y) = -(W dy) q \mu \phi_i \frac{\partial n(x, y)}{\partial x} \quad (\text{J.3})$$

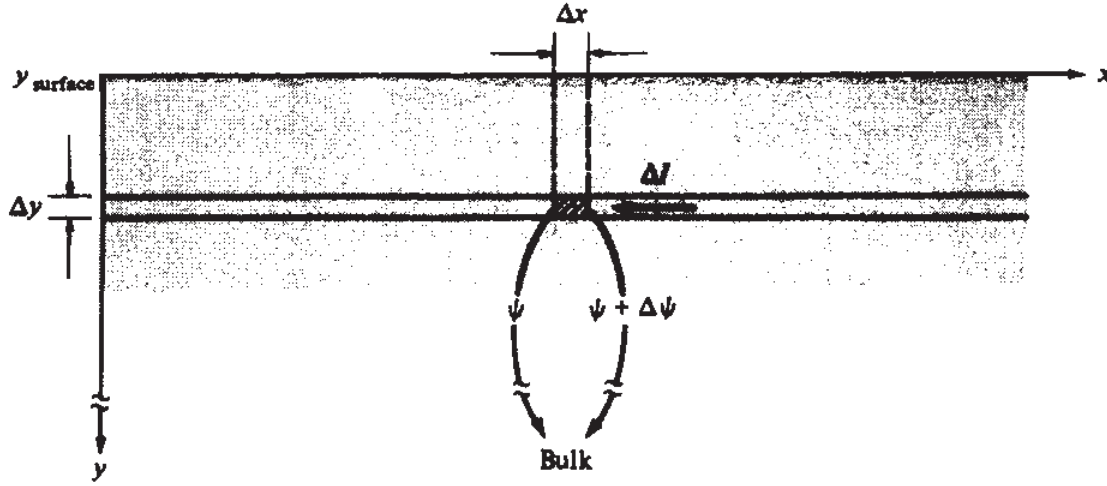


FIGURE J.1

At the source end of the channel, n will be given by (I.1), with $V_{CB} = V_{SB}$, and at the drain end of the channel, n will be given by that relation with $V_{CB} = V_{DB}$. For a position x in the channel, a similar relation can be written with V_{CB} replaced by a quantity $V(x)$:

$$n(x, y) = n_0 e^{[\psi(x, y) - V(x)]/\phi_t} \quad (\text{J.4})$$

where $V(0) = V_{SB}$ and $V(L) = V_{DB}$. At points in the channel between source and drain $V(x)$ takes the value required for (J.4) to provide the correct value of $n(x, y)$. In fact, $V(x)$ can be interpreted as the potential corresponding to the difference between the electron quasi-Fermi level in the inversion layer and the hole quasi-Fermi level in the bulk (Appendix H). This quantity is assumed independent of y in the inversion layer, as discussed in Appendix H.

Differentiating (J.4) with respect to x we obtain

$$\frac{\partial n(x, y)}{\partial x} = \frac{n(x, y)}{\phi_t} \left[\frac{\partial \psi(x, y)}{\partial x} - \frac{dV(x)}{dx} \right] \quad (\text{J.5})$$

Using this in (J.3) gives

$$dI_{\text{diff}}(x, y) = -(W dy) q \mu n(x, y) \left[\frac{\partial \psi(x, y)}{\partial x} - \frac{dV(x)}{dx} \right] \quad (\text{J.6})$$

Using this and (J.2) in (J.1) we see that the $\partial \psi / \partial x$ terms cancel out, resulting in

$$dI_{DS} = (W dy) q \mu n(x, y) \frac{dV(x)}{dx} \quad (\text{J.7})$$

The total drain current can be obtained by integrating over depth from $y = y_{\text{surface}}$ to a point $y = y_c$, below which the electron concentration is negligible. Since V is assumed independent of y as mentioned above, this gives

$$I_{DS} = W\mu \frac{dV(x)}{dx} q \int_{y_{\text{surface}}}^{y_c} n(x, y) dy \quad (\text{J.8})$$

where we have assumed that μ is independent of y . The integral times q can be recognized from (2.5.4) as $-Q'_I$ at position x ; thus

$$I_{DS} = \mu W(-Q'_I) \frac{dV(x)}{dx} \quad (\text{J.9})$$

Integrating over the length of the channel and recognizing $\int_0^L I_{DS} dx$ as $I_{DS}L$, we have

$$I_{DS} = \frac{W}{L} \int_{V_{SB}}^{V_{DB}} \mu(-Q'_I) dV \quad (\text{J.10})$$

Note that the above two relations are of the same form as (4.5.6) and (4.5.7), which were developed by starting from strong-inversion assumptions. However, it is clear from the above development that (J.9) and (J.10) are actually very general. They are valid in all regions of operation, since they include both drift and diffusion effects. The development in Sec. 4.5.1 following (4.5.7) cannot be used here, since Q'_I in (4.5.10) is only valid in strong inversion. A more general expression for Q'_I must be used. Depending on the expression used for Q'_I in conjunction with (J.10), different formulations are possible. These are shown below.

CHARGE SHEET MODEL

In the charge sheet model (Sec. 4.3), Q'_I is known as a function of the surface potential ψ_s [see (4.3.15)]. Thus a change of variables from V to ψ_s is made in (J.10):

$$I_{DS} = \frac{W}{L} \int_{\psi_{s0}}^{\psi_{sL}} \mu(-Q'_I) \frac{dV}{d\psi_s} d\psi_s \quad (\text{J.11})$$

where V is related to ψ_s by an equation analogous to (4.3.18):

$$\psi = V_{GB} - V_{FB} - \gamma \sqrt{\psi_s + \phi_i} e^{(\psi_s - 2\phi_F - V)/\phi_i} \quad (\text{J.12})$$

The quantity $dV/d\psi_s$ can be found from this and substituted in (J.11) to determine the current. This results in a rather lengthy development. Several derivations of the charge sheet model have been proposed.²⁻⁵ The technique we have presented in Sec. 4.3 is the most straightforward.

PAO-SAH MODEL

In this model¹ no simplifying assumptions are made, in the sense that both electrons and holes are allowed to exist in the depletion region. This results, as expected from (I.8), in

$$Q'_t = -qN_A e^{(-2\phi_F - V)/\phi_t} \int_{\psi_c}^{\psi_s} \frac{e^{\psi/\phi_t}}{\mathcal{E}[\psi]} d\psi \quad (\text{J.13})$$

where, consistent with our present convention, we consider the electrons between the surface and a depth beyond which the electron concentration becomes negligible. The potential at that point with respect to the substrate is denoted by ψ_c . A convenient depth is that at which $n = n_i$. From (J.4) and (1.4.1a) we easily see that the corresponding potential is $\psi_c = \phi_F + V$. The quantity $\mathcal{E}[\psi]$ in (J.13) is given in (I.4), with V_{CB} replaced by V . Using (J.13) in (J.10) we obtain

$$I_D = \frac{W}{L} qN_A \int_{V_{SB}}^{V_{DB}} \mu \int_{\psi_c}^{\psi_s} \frac{e^{(\psi - 2\phi_F - V)/\phi_t}}{\mathcal{E}[\psi]} d\psi dV \quad (\text{J.14})$$

This double integral can be evaluated numerically. The required computation time is long, so this formulation is mainly of theoretical interest. The charge sheet model gives nearly the same results and is much simpler. A technique to reduce (J.14) to a single-integral formula (which must still be evaluated numerically) has been proposed.⁶

REFERENCES

1. H. C. Pao and C. T. Sah, "Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors," *Solid-State Electronics*, vol. 10, pp. 927-937, 1966.
2. G. Baccarani, M. Rudan, and G. Spadini, "Analytical i.g.f.e.t. model including drift and diffusion currents," *IEEE Journal on Solid-State and Electron Devices*, vol. 2, pp. 62-68, March 1978.
3. J. R. Brews, "A charge sheet model for the MOSFET," *Solid-State Electronics*, vol. 21, pp. 345-355, 1978.
4. F. Van de Wiele, "A long-channel MOSFET model," *Solid-State Electronics*, vol. 22, pp. 991-997, 1979.
5. J. R. Brews, "Physics of the MOS transistor," chapter 1 in *Silicon Integrated Circuits, Part A*, D. Kahng (editor), Applied Solid-State Science Series, Academic Press, New York, 1981.
6. R. F. Pierret and J. A. Shields, "Simplified long-channel MOSFET theory," *Solid-State Electronics*, vol. 26, pp. 143-147, 1983.

APPENDIX K

RESULTS OF A DETAILED FORMULATION FOR THE DRAIN CURRENT AND DRAIN SMALL-SIGNAL CONDUCTANCE IN THE SATURATION REGION

The results of a detailed analysis of the saturation region are given in this appendix. These results are provided here as examples in order to demonstrate the complexity involved in such analysis. The reader should consult Ref. 1 for the detailed derivations, discussions of the ranges of validity, etc.

The formulation takes into account the drain junction depth and its influence on the inversion layer shape in the vicinity of that junction. The following relation between I_{DS} and V_{DS} in saturation is derived:

$$V_{DS} = V'_{DS} = \frac{qN_A L^2}{2\epsilon_s} \left(1 - \frac{I'_{DS}}{I_{DS}}\right)^2 \left[1 + \frac{2I_{DS}}{qN_A W |v_d|_{\max} d_j} \left(\log \frac{d_j}{d_i} - 1\right)\right] + L\mathcal{E}_1 \left(1 - \frac{I'_{DS}}{I_{DS}}\right) \quad (\text{K.1})$$

where $|v_d|_{\max}$ is the maximum (saturated) velocity magnitude of carriers in the pinchoff region, \mathcal{E}_1 the approximate value of field intensity at which velocity saturation is reached (roughly) $|v_d|_{\max}/\mu$; see Fig. 6.22 and the associated discussion); d_j is the drain junction depth, and d_i is the average thickness of the inversion layer (assumed much smaller than d_j and taken to be typically 100 Å). Differentiation of this equation gives the following value for g_{sd} (Sec. 8.2):

$$g_{sd} = \left\{ \frac{qN_A L^2}{2\epsilon_s} \left(1 - \frac{I'_{DS}}{I_{DS}}\right) \left[2 \frac{I'_{DS}}{I_{DS}^2} + \left(1 + \frac{I'_{DS}}{I_{DS}}\right) \frac{2}{qN_A W v_{\max} d_j} \left(\log \frac{d_j}{d_i} - 1\right) \right] + L\mathcal{E}_1 \frac{I'_{DS}}{I_{DS}^2} \right\}^{-1} \quad (\text{K.2})$$

BIBLIOGRAPHY

G. Merckel, J. Borel, and N. Z. Cupcea, "An accurate large-signal MOS transistor model for use in computer-aided design," *IEEE Transactions on Electron Devices*, vol. Ed-19, pp. 681-690, May 1972.

APPENDIX L

EVALUATION OF THE INTRINSIC TRANSIENT SOURCE AND DRAIN CURRENTS

We present here a proof¹ of certain statements made in Sec. 7.3 for the intrinsic part of a MOS transistor. The continuity equation is, from (7.7.5),

$$\frac{\partial i(x, t)}{\partial x} = W \frac{\partial q'_I(x, t)}{\partial t} \quad (\text{L.1})$$

Integrating with respect to distance from the source to a point x in the channel, we obtain

$$i(x, t) - i(0, t) = W \int_0^x \frac{\partial q'_I(\hat{x}, t)}{\partial t} d\hat{x} \quad (\text{L.2})$$

where \hat{x} is a dummy variable of integration. Recognizing $i(0, t)$ as $-i_S(t)$ (see the direction of i in Fig. 7.13), and using (7.7.6b) for $i(x, t)$ we obtain, from (L.2),

$$i_S(t) = \mu W q'_I(x, t) \frac{\partial v_{CB}(x, t)}{\partial x} + W \int_0^x \frac{\partial q'_I(\hat{x}, t)}{\partial t} d\hat{x} \quad (\text{L.3})$$

Multiplying both sides by dx , integrating from $x = 0$ to $x = L$, and dividing both sides by L gives

$$i_s(t) = \frac{W}{L} \int_0^L \mu q'_l(x, t) \frac{\partial v_{CB}(x, t)}{\partial x} dx + \frac{W}{L} \int_0^L \int_0^x \frac{\partial q'_l(\hat{x}, t)}{\partial t} d\hat{x} dx \quad (\text{L.4})$$

Interchanging the order of integration and differentiation in the second term of the right-hand side, we obtain

$$i_s(t) = \frac{W}{L} \int_0^L \mu q'_l(x, t) \frac{\partial v_{CB}(x, t)}{\partial x} dx + \frac{d}{dt} \left[\frac{W}{L} \int_0^L \int_0^x q'_l(\hat{x}, t) d\hat{x} dx \right] \quad (\text{L.5})$$

The double integral is of the form $\int_0^L G(x) dx$, where $G(x) = \int_0^x q'_l(\hat{x}, t) d\hat{x}$. Applying integration by parts to $\int_0^L G(x) dx$, with G and x the two variables involved, we can write (L.5) as follows:

$$i_s(t) = -\frac{W}{L} \int_0^L \mu [-q'_l(x, t)] \frac{\partial v_{CB}(x, t)}{\partial x} dx + \frac{d}{dt} \left[W \int_0^L \left(1 - \frac{x}{L} \right) q'_l(x, t) dx \right] \quad (\text{L.6})$$

Comparing the first term to the right-hand side of (4.5.7), we see that, in quasi-static operation, this term will produce the same current expressions as in Sec. 4.5, only with a minus sign and with the terminal voltages as functions of time. We can then write (L.6) as

$$i_s(t) = -i_T(t) + \frac{dq_S}{dt} \quad (\text{L.7})$$

where $i_T(t)$ will be of the form of (7.3.12), and

$$q_S = W \int_0^L \left(1 - \frac{x}{L} \right) q'_l dx \quad (\text{L.8})$$

To find $i_D(t) = i(L, t)$ in quasi-static operation, we use $x = L$ in (L.2) and substitute in it (L.7) and (L.8), which gives

$$i_D(t) = i_T(t) + \frac{dq_D}{dt} \quad (\text{L.9})$$

with

$$q_D = W \int_0^L \frac{x}{L} q'_l dx \quad (\text{L.10})$$

Since quasi-static operation is assumed, (L.8) and (L.10) can be evaluated by using the dc charge per unit area Q'_l instead of q'_l . The resulting quantities have been denoted by Q_S and Q_D , respectively, in (7.3.9b) and (7.3.9a). The detailed evaluation of Q_S and Q_D is considered in Sec. 7.4.

A NOTE ON q_D AND q_S

If desired, $q_D(t)$ can be defined as $\int_{-\infty}^t [i_D(\tau) - i_T(\tau)] d\tau = \int_{-\infty}^t i_{DA}(\tau) d\tau$, where τ is a dummy variable of integration and $q_S(t)$ as $\int_{-\infty}^t [i_S(\tau) + i_T(\tau)] d\tau = \int_{-\infty}^t i_{SA}(\tau) d\tau$ (i_{DA} and i_{SA} are the "charging" currents discussed in Sec. 7.3). These definitions imply (7.3.6), and, by integrating (7.3.5), we see that they satisfy (7.3.8) (assuming that the channel was "empty" at $t = -\infty$). With these definitions, $q_D(t)$ can be interpreted as the part contributed to $q_I(t)$ by the deviation of i_D from i_T , from the "beginning of time" to the instant t . Similarly, $q_S(t)$ can be interpreted as the part contributed to $q_I(t)$ by the corresponding deviation of i_S from $-i_T(t)$. Although the above definitions may be elegant, they can confuse the issue: q_D and q_S cannot be found stored in any particular place, and neither can they be associated with specific charges that have been flowing by themselves through any terminal. This is because $i_{DA}(t)$ and $i_{SA}(t)$ are mere artifacts, each being the difference between *real* terminal currents and the value such terminal currents *would* have if one attempted to predict them by dc theory [see (7.3.4)]. Also, the above definitions for q_D and q_S tend to assign specific significance to the *values* of these fictitious charges, when all that matters is their *derivatives* [see (7.3.6)]. Assume q_D and q_S defined as above are given by two functions f_D and f_S , respectively [where f_D and f_S are the same functions as in (7.3.10)], giving the correct values for i_{DA} and i_{SA} from (7.3.6). The quantities q_D and q_S satisfy (7.3.8), as already noted. Now let us redefine q_D and q_S as $f_D + K$ and $f_S + L$, respectively, where K and L are arbitrary constants. Then (7.3.6) will still predict the correct values of i_{DA} and i_{SA} . Equation (7.3.8) will, in general, not hold now, but this is of *no* consequence and does *not* mean that charge conservation is violated. This is because, since i_{DA} and i_{SA} are still predicted correctly, they still satisfy (7.3.5) and (7.3.3). Integrating (7.3.3), one sees that the total charge that entered the device through source and drain up to time t is equal to $q_I(t)$, and thus the charge is conserved.

REFERENCE

1. D. E. Ward, "Charge-based modeling of capacitance in MOS transistors," Technical Report G201-11, Integrated Circuits Laboratory, Stanford University, June 1981.

APPENDIX M

CHARGES FOR THE ACCURATE STRONG-INVERSION MODEL

Using the charges per unit area as they correspond to the complete strong-inversion model of Sec. 4.5.1 [(4.5.9) and (4.5.10)], and proceeding as in Sec. 7.4.2, we obtain for the nonsaturation region¹

$$Q_B = -WLC'_{ox} \frac{\frac{2}{3}\gamma U_G(U_D^{2/3} - U_S^{2/3}) - \frac{1}{2}\gamma^2(U_D^2 - U_S^2) - \frac{2}{5}\gamma(U_D^{5/2} - U_S^{5/2})}{\hat{g}(U_G, U_D, U_S)} \quad (M.1)$$

$$Q_I = -WLC'_{ox} \left[\frac{U_G^2(U_D - U_S) - \frac{4}{3}\gamma U_G(U_D^{3/2} - U_S^{3/2})}{\hat{g}(U_G, U_D, U_S)} + \frac{(U_G - \gamma^2/2)(U_D^2 - U_S^2) - \frac{4}{3}\gamma(U_D^{5/2} - U_S^{5/2}) - \frac{1}{3}(U_D^3 - U_S^3)}{\hat{g}(U_G, U_D, U_S)} \right] \quad (M.2)$$

$$Q_G = WLC'_{ox} \left[U_G - \frac{\frac{1}{2}U_G(U_D^2 - U_S^2) - \frac{2}{5}\gamma(U_D^{5/2} - U_S^{5/2}) - \frac{1}{3}(U_D^3 - U_S^3)}{\hat{g}(U_G, U_D, U_S)} \right] - Q_o \quad (M.3)$$

with

$$\hat{g}(U_G, U_D, U_S) = U_G(U_D - U_S) - \frac{2}{3}\gamma(U_D^{2/3} - U_S^{2/3}) - \frac{1}{2}(U_D^2 - U_S^2) \quad (\text{M.4})$$

where

$$U_G = V_{GB} - V_{FB} \quad (\text{M.5})$$

$$U_D = V_{DB} + \phi_0 \quad (\text{M.6})$$

$$U_S = V_{SB} + \phi_0 \quad (\text{M.7})$$

In saturation, of course, V_{DB} should be replaced by V_p as given by (4.5.11).

The above expressions have the problem that the numerator and denominator become zero as $U_D - U_S$ goes to zero. This problem can be bypassed if one factors out the quantity $U_D^{1/2} - U_S^{1/2}$ from numerator and denominator.¹

Sometimes simpler, empirical expressions are used in conjunction with this model. In choosing such expressions, one should seek to preserve the symmetry inherent in the model. Also, the expressions should be such that charge saturation is attained at the same voltage combinations as drain current saturation. For example, an empirical expression for Q_I in nonsaturation that satisfies these requirements is²

$$Q_I = -C'_{ox}WL \frac{2}{3} \left(V_1 + V_2 - \frac{V_1 V_2}{V_1 + V_2} \right) \quad (\text{M.8})$$

where

$$V_1 = V_{GS} - V_T(V_{SB}) \quad (\text{M.9})$$

$$V_2 = V_{GD} - V_T(V_{DB}) \quad (\text{M.10})$$

and

$$V_T(V_{SB}) = V_{FB} + \phi_0 + \gamma\sqrt{\phi_0 + V_{SB}} \quad (\text{M.11})$$

$$V_T(V_{DB}) = V_{FB} + \phi_0 + \gamma\sqrt{\phi_0 + V_{DB}} \quad (\text{M.12})$$

By using two different thresholds as corresponding to the source and drain biases, total symmetry is ensured for Q_I with respect to V_S and V_D . In saturation at the drain, Q_I can be obtained by using $V_{GD} = V_T(V_{DB})$, i.e., $V_2 = 0$; it is easy to see that this results in (7.4.27).

The Q_D and Q_S expressions resulting from exact calculations for the complete strong-inversion model of Sec. 4.5.1 are *extremely* complicated. Thus for these two charges empirical expressions are used. Two such expressions are¹

$$Q_D \approx \frac{2}{5} Q_{I,\text{sat}} + \frac{7}{10} (Q_I - Q_{I,\text{sat}}) \quad (\text{M.13})$$

$$Q_S \approx \frac{3}{5} Q_{I,\text{sat}} + \frac{3}{10} (Q_I - Q_{I,\text{sat}}) \quad (\text{M.14})$$

where Q_I can be modeled empirically as discussed above. It is easy to see that, in saturation, (M.13) and (M.14) reduce to (7.4.28) and (7.4.29), respectively. Note that the above empirical expressions for Q_D and Q_S can be used in conjunction with the approximate model as well in lieu of (7.4.19) and (7.4.20).

Attempting to calculate capacitances (Sec. 8.3) from the charges corresponding to the accurate strong-inversion model results in very complicated expressions.³ One thus resorts to empirical approximations. For C_{gs} and C_{gd} the following approximations have been proposed:

$$C_{gs} = \frac{2}{3} C_{ox} \left[1 - \left(\frac{V_2}{V_1 + V_2} \right)^2 \right] \quad (\text{M.15})$$

$$C_{gd} = \frac{2}{3} C_{ox} \left[1 - \left(\frac{V_1}{V_1 + V_2} \right)^2 \right] \quad (\text{M.16})$$

with V_1 and V_2 defined in (M.9) to (M.12). The above expressions exhibit the symmetry characteristic of the complete strong-inversion model. In saturation at the drain, we have $V_2 = 0$; this gives $C_{gs} = \frac{2}{3} C_{ox}$ and $C_{gd} = 0$, in agreement with (8.3.24) and (8.3.26).

REFERENCES

1. D. E. Ward, "Charge-based modeling of capacitances in MOS transistors," Technical Report G201-11, Integrated Circuits Laboratory, Stanford University, June, 1981.
2. F. M. Klaassen, "A MOS model for computer-aided design," *Philips Research Reports*, vol. 31, pp. 71-83, 1976.
3. R. S. C. Cobbold, *Theory and Application of Field-Effect Transistors*, Wiley-Interscience, New York, 1970.

APPENDIX N

QUANTITIES USED IN THE DERIVATION OF THE NON-QUASI-STATIC Y-PARAMETER MODEL

Expressions for y parameters, including non-quasi-static effects, have been given in (9.4.58) to (9.4.60). In these expressions, the numerators $N_{kl}(\omega)$ (where $k, l = d, g, b$) are of the form

$$N_{kl}(\omega) = n_{kl0} + (j\omega)n_{kl1} + (j\omega)^2n_{kl2} + \dots$$

and the denominator $D(\omega)$ is of the form

$$D(\omega) = d_0 + (j\omega)d_1 + (j\omega)^2d_2 + \dots$$

The coefficients in the above relations can be found as discussed in Sec. 9.4.2; the results are given below. Some of these results apply to the three-terminal transistor with $\alpha_1 = 1$;¹ all the results apply to the four-terminal transistor with $\alpha_1 = 1$.²

$$n_{dd0} = \frac{W}{L} \mu C'_{ox} (V_{GS} - V_T) \eta = g_{sd}$$

$$n_{dg0} = \frac{W}{L} \mu C'_{ox} \frac{V_{GS} - V_T}{\alpha_1} (1 - \eta) = g_m$$

$$n_{db0} = (\alpha_1 - 1) n_{dg0} = g_{mb}$$

$$n_{gd0} = n_{gg0} = n_{bd0} = n_{gb0} = n_{bg0} = n_{bb0} = 0$$

$$n_{dd1} = \frac{2}{3} \alpha_1 C_{\text{ox}} \frac{\eta(2 + \eta)}{(1 + \eta)^2}$$

$$n_{dg1} = n_{gd1} = -\frac{2}{3} C_{\text{ox}} \frac{\eta(2 + \eta)}{(1 + \eta)^2}$$

$$n_{db1} = (\alpha_1 - 1) n_{dg1}$$

$$n_{gg1} = C_{\text{ox}} \left[\frac{2}{3\alpha_1} \frac{1 + 4\eta + \eta^2}{(1 + \eta)^2} + \frac{\alpha_1 - 1}{\alpha_1} \right]$$

$$n_{gb1} = n_{bg1} = -\frac{\alpha_1 - 1}{3\alpha_1} C_{\text{ox}} \left(\frac{1 - \eta}{1 + \eta} \right)^2$$

$$n_{bd1} = (\alpha_1 - 1) n_{gd1}$$

$$n_{bb1} = C_{\text{ox}} \left[\frac{\alpha_1 - 1}{\alpha_1} + \frac{2(\alpha_1 - 1)^2}{3\alpha_1} \frac{1 + 4\eta + \eta^2}{(1 + \eta)^2} \right]$$

$$n_{dd2} = \frac{C_{\text{ox}}}{\omega_o} \frac{2}{45} \alpha_1 \frac{\eta(5 + 8\eta + 2\eta^2)}{(1 + \eta)^4}$$

$$n_{dg2} = n_{gd2} = -\frac{C_{\text{ox}}}{\omega_o} \frac{2}{45} \frac{\eta(5 + 8\eta + 2\eta^2)}{(1 + \eta)^4}$$

$$n_{db2} = (\alpha_1 - 1) n_{dg2}$$

$$n_{gg2} = \frac{C_{\text{ox}}}{\omega_o} \left[\frac{2}{45} \frac{1}{\alpha_1} \frac{2 + 11\eta + 2\eta^2}{(1 + \eta)^3} + \frac{4}{15} \frac{\alpha_1 - 1}{\alpha_1} \frac{1 + 3\eta + \eta^2}{(1 + \eta)^3} \right]$$

$$n_{gb2} = n_{bg2} = -\frac{C_{\text{ox}}}{\omega_o} \frac{2}{45} \frac{\alpha_1 - 1}{\alpha_1} \frac{4 + 7\eta + 4\eta^2}{(1 + \eta)^3}$$

$$n_{bd2} = (\alpha_1 - 1) n_{gd2}$$

$$n_{bb2} = \frac{C_{\text{ox}}}{\omega_o} \left[\frac{4}{15} \frac{\alpha_1 - 1}{\alpha_1} \frac{1 + 3\eta + \eta^2}{(1 + \eta)^3} + \frac{2}{45} \frac{(\alpha_1 - 1)^2}{\alpha_1} \frac{2 + 11\eta + 2\eta^2}{(1 + \eta)^3} \right]$$

$$d_0 = 1$$

$$d_1 = \frac{4}{15} \frac{1}{\omega_o} \frac{1 + 3\eta + \eta^2}{(1 + \eta)^3}$$

$$d_2 = \frac{1}{45} \frac{1}{\omega_o^2} \frac{1 + 4\eta + \eta^2}{(1 + \eta)^4}$$

where

$$C_{ox} = C'_{ox} WL$$

$$\omega_o = \frac{\mu(V_{GS} - V_T)}{\alpha L^2}$$

$$\eta = \begin{cases} 1 - \frac{V_{DS}}{V'_{DS}}, & V_{DS} \leq V'_{DS} \\ 0, & V_{DS} \geq V'_{DS} \end{cases}$$

and, according to the simplifying assumptions stated in the beginning of Sec. 9.4.2,

$$\alpha_1 = 1 + \frac{\gamma}{2\sqrt{\phi_0 + V_{SB}}}$$

As explained in Sec. 9.4, the accuracy of the expressions containing α_1 can be improved if α_1 is replaced by an empirical parameter which, in general, would not be the same for all expressions; a related discussion can be found in Sec. 8.3.

REFERENCES

1. J. J. Paulos and D. Antoniadis, "Limitations of quasi-static models for the MOS transistor," *IEEE Electron Device Letters*, vol. ED-4, pp. 221-224, July 1983.
2. M. Bagheri and Y. Tsividis, "A small-signal non-quasi-static model for the four-terminal MOSFET valid in all regions of operation," *IEEE Transactions on Electron Devices*, vol. ED-32, pp. 2383-2391, November 1985.

A

Abrupt (step) junctions, 26–27
 Acceptors, 5–7
 Accumulation
 gate-substrate voltage and, 59, 60
 intrinsic capacitances and, 402, 403
 quasi-static operation and, 336
 Admittance matrix representation, 462
 Atalla, M. M., 34
 Avalanche effect, 191, 286

B

Back gate, 41
 Barrier lowering, 257–259, 264–265, 267
 Benchmark tests, 534–543
 Body (substrate), 34, 35
 Body (substrate) effect
 in four-terminal MOS transistors, 61
 introduced, 40–41
 in three-terminal MOS structures,
 101–103
 Body (substrate) terminal, 51, 52
 Body effect coefficient, 68, 69, 102,
 105, 192
 Body-referenced versus source-referenced
 modeling, 179–181
 Boltzmann limit, 379
 Boltzmann's constant, 412
 Bonds between atoms, 2
 Breakdown, 191–192
 Built-in potentials, 27
 Bulk mobility, 12–14, 236, 237
 Bulk punchthrough, 278–279
 Buried channel devices, 226, 228, 241

Buried-channel enhancement-mode
 *p*MOS, 241
 Buried-channel mode, 236n

C

Capacitances
 in the complete quasi-static model
 description of effects, 440–445
 evaluation of, 451–458, 459–460
 extrinsic, 405–409, 410
 gate, 406–408
 intrinsic
 in accumulation, 402, 403
 in depletion, 400
 general charge sheet model, 400,
 401, 402
 intrinsic transition (cutoff)
 frequency, 397–398
 introduced, 385–390
 in moderate inversion, 398–399
 in nonsaturation, 393–395
 in saturation, 395–397
 small-dimension effects, 402–405
 in strong inversion, 390–393
 in weak inversion, 399–400
 junction, 408–409
 reverse-biased *pn* junctions, 31–33
 source-drain proximity, 409
 two-terminal MOS structures, 79–86
 Capacitors, MOS; *see* Two-terminal MOS
 structures
 Carrier velocity saturation, 280–286
 Carriers
 intrinsic carrier concentration, 3
 majority and minority, 4, 5

- Channel; *see also* Small-dimension effects
 - defined, 35
 - effective channel width, 273–275
 - gradual channel approximation, 127–128
 - narrow-channel devices, 270–277
 - short-channel devices, 257–269
 - Channel breakdown, 191
 - Channel-length modulation (CLM), 249–257
 - Charge balance, 56–58, 67
 - Charge density, 7–8, 566–571
 - Charge pumping, 346
 - Charge sharing, 259
 - Charge sheet approximation, 65
 - Charge sheet models
 - complete charge sheet model, 131–140
 - intrinsic capacitances and, 400, 401, 402
 - models based on quasi-Fermi potentials, 146, 594–597
 - quasi-static operation and, 333–335
 - simplified charge sheet models, 140–146
 - summarized, 169–170
 - Circuit simulation, models for
 - accuracy, 528–529
 - benchmark tests, 534–543
 - combining several effects into one physical model, 516–518
 - considerations and choices
 - considerations related to computer implementation, 533–534
 - general considerations, 530–533
 - nontechnical considerations, 543
 - introduced, 513–514
 - parameter extraction, 519–528
 - properties of good models, 529–530
 - types of models, 514–515
 - CLM (channel-length modulation), 249–257
 - CMOS (Complementary MOS)
 - technology, 36–38
 - Compact models, 514–515
 - Compensated buried-channel devices, 241
 - Complementary MOS (CMOS)
 - technology, 36–38
 - Computer simulation; *see* Circuit simulation, models for
 - Conductance; *see also* Transconductance
 - defined, 13
 - output, 376–378, 535, 537–538, 598–599
 - source-drain, 363, 370–375
 - Conduction
 - diffusion, 15–18
 - drift, 9–15
 - transit time, 8–9, 13, 17–18
 - Conductivity, 13
 - Constant current threshold voltage, 196
 - Constant-field scaling, 291–294
 - Constant-voltage scaling, 294
 - Contact potentials, 18–25, 561–562
 - Contact windows, 36
 - Continuity equation, 347–348
 - Corner frequency, 413
 - Coulomb scattering, 181, 182
 - Crystal lattice, 2
 - Cutoff (transition) frequency, 500–502
- ## D
- Degenerate semiconductors, 4
 - Depletion
 - gate-substrate voltage and, 59–62
 - intrinsic capacitances and, 400
 - quasi-static operation and, 335–336
 - Depletion approximation, 27, 65
 - Depletion-mode transistors
 - n*MOS
 - approximate strong-inversion model for *n*-implant-*p*-substrate devices, 238
 - charges and threshold voltage, 227–234
 - introduced, 194–195

need for n -type implants, 225–227
 transistor operation, 234–240
*p*MOS, 194–195
 Depletion region, 27, 40–41
 DIBL (drain-induced barrier lowering),
 264–265, 267
 Dielectric constant, 8
 Diffusion, 15–18, 35
 Diffusion constant, 16
 Direct tunneling, 289
 Donors, 4–7
 Doping, 3–7, 298–301
 Drain, 35, 36
 Drain conductance, 376–378, 535,
 537–538, 598–599
 Drain current, 594–597, 598–599
 Drain-induced barrier lowering (DIBL),
 264–265, 267
 Drain series resistance, 296–298
 Drift, 9–15
 Drift velocity, 10
 Dynamic operation of MOS transistors;
 see Large-signal dynamic
 operation of MOS transistors

E

Ebers-Moll equations, 172
 Effective channel width, 273–275
 Effective interface charge, 54
 Effective mobility, 181–189
 Effective threshold voltage
 introduced, 257
 in narrow-channel devices, 270–277
 in short-channel devices, 257–269
 summarized, 277
 Einstein relationship, 16
 EKV model, 176–178
 Electric fields, 566–571
 Electrons
 free, 2–7
 valence, 2

Electrostatics in one dimension, 566–571
 Empirical models, 514–515
 Energy bands
 in semiconductors, 557–561
 in three-terminal MOS structures,
 587–590
 in two-terminal MOS structures,
 574–577
 Enhancement-mode transistors
 *n*MOS
 charges and threshold voltages,
 212–219
 drain-to-source current model for
 strong inversion, 219–222
 introduced, 194–195
 preliminaries, 208–212
 simplified model for strong
 inversion, 222–224
 weak inversion, 225
 *p*MOS
 buried-channel enhancement-mode
 *p*MOS, 241
 introduced, 194–195
 surface-channel enhancement-mode
 *p*MOS, 240–241
 Epitaxy, 38
 Equivalent input noise resistance, 420
 Equivalent noise voltage, 420
 Extrapolated threshold voltage, 72–73,
 108, 159, 160–161, 169n, 196–197
 Extrinsic capacitances, 405–409, 410
 Extrinsic resistances, 409–410
 Extrinsic semiconductors, 3–7, 558

F

Fast surface-state (interface trap) charge, 54
 Fermi-Dirac statistics, 558
 Fermi potentials, 19–21, 561
 Field oxide, 270
 Flat-band condition, 575
 Flat-band voltage, 51–56, 59, 214n

Flicker noise, 413, 422–424, 539
 Forward bias, 26, 29
 Forward saturation, 154–155
 Free electrons, 2–7
 Fringing fields, 270, 272

G

Gate
 back, 41
 introduced, 35, 36
 polysilicon gate depletion, 299–300
 Gate capacitances, 406–408
 Gate oxide effective thickness, 299
 Gate resistance, 498–500
 Gate-source extrapolated threshold voltage, 159, 160–161
 Gate-substrate voltage
 effect on surface condition
 accumulation, 59, 60
 depletion and inversion, 59–62
 flat-band condition, 59
 general analysis, 62–64, 67
 introduced, 58
 extrapolated threshold voltage, 108
 Gate terminal, 51, 52
 Gate transconductance, 361, 367–368
 Generalized scaling, 294
 Gradual channel approximation, 127–128

H

Heil, O., 34
 High-frequency modeling; *see* Small-signal modeling for high frequencies
 High-frequency noise, 492–495
 High-frequency transadmittance, 538–539
 Holes, 2–7
 Hot carrier effects, 286–290, 424

I

IGFETs (insulated-gate field-effect transistors); *see* MOS (metal oxide semiconductor) transistors
 Impact ionization, 191, 286
 Impurities, 3–7
 Induced gate noise, 492–495
 Insulated-gate field-effect transistors (IGFETs); *see* MOS (metal oxide semiconductor) transistors
 Interface trap (fast surface-state) charge, 54
 Interpolation models, for four-terminal MOS structures, 176–178
 Intrinsic capacitances
 in accumulation, 402, 403
 in depletion, 400
 general charge sheet model, 400, 401, 402
 intrinsic transition (cutoff) frequency, 397–398
 introduced, 385–390
 in moderate inversion, 398–399
 in nonsaturation, 393–395
 in saturation, 395–397
 small-dimension effects, 402–405
 in strong inversion, 390–393
 in weak inversion, 399–400
 Intrinsic carrier concentration, 3
 Intrinsic cutoff (transition) frequency, 397–398
 Intrinsic semiconductors, 2–3, 557–558
 Intrinsic transition (cutoff) frequency, 397–398
 Inversion; *see also* Moderate inversion; Strong inversion; Weak inversion
 defined, 39–40
 in four-terminal MOS structures
 moderate inversion, 175–176
 regions in terms of terminal voltages, 147–150
 regions of operation, 128–130

strong inversion; *see* Strong inversion, in four-terminal MOS structures
 weak inversion, 170–175
 gate-substrate voltage and, 59–62
 in three-terminal MOS structures
 approximate limits of regions, 103–106
 moderate inversion, 113, 122, 586
 strong inversion, 106–109, 121
 weak inversion, 109–113, 121
 in two-terminal MOS structures
 general relations and regions, 64–71, 582–583
 moderate inversion, 78–79, 87, 584–585
 strong inversion, 71–74, 87
 summary of properties, 87
 weak inversion, 74–78, 87
 Inversion layer, 40
 Ion implantation
 in depletion *n*MOS transistors
 charges and threshold voltage, 227–234
 need for *n*-type implants, 225–227
 transistor operation, 234–240
 in enhancement *n*MOS transistors
 charges and threshold voltages, 212–219
 drain-to-source current model for
 strong inversion, 219–222
 preliminaries, 208–212
 simplified model for strong inversion, 222–224
 weak inversion, 225
 in enhancement *p*MOS transistors
 buried-channel enhancement-mode *p*MOS, 241
 surface-channel enhancement-mode *p*MOS, 240–241
 introduced, 207–208

J

Johnson (Nyquist, thermal) noise, 412, 414, 539
 Junction breakdown, 191
 Junction capacitances, 408–409

K

Kahng, D., 34

L

Large-signal dynamic operation of MOS transistors
 introduced, 312–313
 non-quasi-static modeling
 analysis, 348–354
 continuity equation, 347–348
 introduced, 347
 quasi-static operation
 accumulation, 336
 depletion, 335–336
 general charge sheet model, 333–335
 introduced, 313–317, 325
 limitations, 340–346
 moderate inversion, 331–332
 plots of charges versus V_{GS} , 336–337
 strong inversion, 325–331, 603–605
 terminal currents in, 317–324, 600–602
 transit time under dc conditions, 339–340, 341
 use of charges in evaluating terminal currents, 337–339
 weak inversion, 332–333
 Lateral diffusion, 35
 Lightly doped drain (LDD) structure, 289–290
 Lilienfeld, J. E., 34

Linear regression, 520
 Local oxidation of silicon (LOCOS) CMOS process, 36–38, 270–275, 276
 Low-frequency modeling; *see* Small-signal modeling for low and medium frequencies

M

Majority carriers, 4, 5
 Mathiessen rule for mobility, 184n
 Maximum frequency of oscillation (unity power gain frequency), 502–504
 Maxwell-Boltzmann statistics, 558–559
 Medium-frequency modeling; *see* Small-signal modeling for low and medium frequencies
 Metal oxide semiconductor transistors; *see* MOS (metal oxide semiconductor) transistors
 Minority carriers, 4, 5
 Mobile ionic charge, 54
 Mobility
 bulk, 12–14, 236, 237
 effective, 181–189
 Mathiessen rule for, 184n
 surface, 132, 181–182, 237
 Moderate inversion
 in four-terminal MOS transistors, 44–45, 46, 175–176
 intrinsic capacitances and, 398–399
 in low-frequency model for the intrinsic part, 380–381
 quasi-static operation and, 331–332
 in three-terminal MOS structures, 113, 122, 586
 in two-terminal MOS structures, 78–79, 87, 584–585
 MOS (metal oxide semiconductor) transistors; *see also* Small-dimension effects
 breakdown, 191–192

charge sheet models
 complete charge sheet model, 131–140
 drain current formulation and, 596
 models based on quasi-Fermi potentials, 146, 594–597
 simplified charge sheet models, 140–146
 summarized, 169–170
 depletion-mode; *see* Depletion-mode transistors
 effective mobility, 181–189
 enhancement-mode; *see* Enhancement-mode transistors
 interpolation models, 176–178
 introduced, 125–128
 inversion
 moderate inversion, 175–176
 regions in terms of terminal voltages, 147–150
 regions of operation, 128–130
 strong inversion; *see* Strong inversion, in four-terminal MOS structures
 weak inversion, 170–175
 model parameter values, accuracy, and comparisons, 195–197
*n*MOS; *see n*-channel MOS (*n*MOS) transistors
*p*MOS; *see p*-channel MOS (*p*MOS) transistors
 regions of operation, 128–130
 source-referenced versus body-referenced modeling, 179–181
 surface mobility, 132, 181–182
 temperature effects, 189–191
 MOS capacitors; *see* Two-terminal MOS structures
 MOS field-effect transistors (MOSFETs); *see* MOS (metal oxide semiconductor) transistors

N

n-channel MOS (*n*MOS) transistors
 depletion-mode
 approximate strong-inversion model
 for *n*-implant-*p*-substrate devices,
 238
 charges and threshold voltage,
 227–234
 introduced, 194–195
 need for *n*-type implants, 225–227
 transistor operation, 234–240
 enhancement-mode
 charges and threshold voltages,
 212–219
 drain-to-source current model for
 strong inversion, 219–222
 introduced, 194–195
 preliminaries, 208–212
 simplified model for strong
 inversion, 222–224
 weak inversion, 225
 introduced, 35–36, 39–41, 126–127
n-implant-*p*-substrate devices, approximate
 strong-inversion model for, 238
n type semiconductors, 4
n-well processes, 36
 Narrow-channel devices, 270–277
*n*MOS transistors; *see n*-channel MOS
 (*n*MOS) transistors
 Noise
 equivalent-circuit model, 425–426
 equivalent input noise resistance, 420
 equivalent noise voltage, 420
 flicker noise, 413, 422–424, 539
 high-frequency, 492–495
 induced gate noise, 492–495
 introduced, 410–414
 random telegraph noise, 424–425
 shot noise, 421
 small-dimension effects, 424–425

 thermal (Johnson, Nyquist) noise, 412,
 414, 539
 white noise, 412, 413, 414–422
 Non-quasi-static models
 analysis, 348–354
 continuity equation, 347–348
 introduced, 347, 467–468
 model comparisons, 491–492
 other approximations and higher-order
 models, 490–491
 strong-inversion model; *see* Non-quasi-
 static strong-inversion model
 Non-quasi-static strong-inversion model
 complex exponential excitation,
 475–489, 606–608
 dc (bias) excitation, 469–471
 introduced, 469
 small-signal excitation, 472–475
 time-varying excitation, 471–472
 Nonsaturation region, 40, 41, 44–46,
 129n, 129
 Nyquist (Johnson, thermal) noise, 412,
 414, 539

O

Output conductance, 376–378, 535,
 537–538
 Oxide, 34, 35
 Oxide breakdown, 192
 Oxide effects, 298–301
 Oxide fixed charge, 54
 Oxide trapped charge, 54

P

p-channel MOS (*p*MOS) transistors
 depletion-mode, 194–195

- enhancement-mode
 - buried-channel enhancement-mode *p*MOS, 241
 - introduced, 194–195
 - surface-channel enhancement-mode *p*MOS, 240–241
 - introduced, 192–194
- p* type semiconductors, 5
- Pao-Sah model, 597
- passivation layers, 36
- Pentode region, 155n
- Permittivity, 8
- Phonon scattering, 181
- Phonons, 181
- Physical models, 514–515
- Pinchoff voltage, 117–122, 154–155, 228n
- p*MOS transistors; *see p*-channel MOS (*p*MOS) transistors
- pn* junctions
 - charge density, electric field, and potential in, 572–573
 - energy band diagram and potential under reverse bias, 564–565
 - introduced, 26–33
- Poisson's equation, 7–8
- Polysilicon (poly), 35
- Polysilicon gate depletion, 299–300
- Potential balance, 56–58, 67
- Potentials
 - built-in, 27
 - contact, 18–25, 561–562
 - and electric fields, 566–571
 - Fermi, 19–21, 561
 - quasi-Fermi, 146, 334n, 594–597
 - surface, 56–58
- Power spectral density, 411
- Punchthrough, 192n, 210, 277–279

Q

- Quantum mechanical effects, 300
- Quasi-constant-voltage scaling, 294

- Quasi-Fermi levels, 116n, 135n, 152n, 562–565
- Quasi-Fermi potentials, 146, 334n, 594–597
- Quasi-static model, complete; *see also* Quasi-static operation
 - description capacitance effects, 440–445
 - equivalent circuit topologies, 445–451
 - evaluation of capacitances, 451–458, 459–460
 - frequency region of validity, 458, 460
- Quasi-static operation; *see also* Quasi-static model, complete
 - accumulation, 336
 - depletion, 335–336
 - general charge sheet model, 333–335
 - introduced, 313–317, 325
 - limitations, 340–346
 - moderate inversion, 331–332
 - plots of charges versus V_{GS} , 336–337
 - strong inversion, 325–331, 603–605
 - terminal currents in, 317–324, 600–602
 - transit time under dc conditions, 339–340, 341
 - use of charges in evaluating terminal currents, 337–339
 - weak inversion, 332–333

R

- Radio-frequency modeling; *see* RF modeling
- Random telegraph noise, 424–425
- Recombination, 2, 4
- Resistances
 - drain series, 296–298
 - extrinsic, 409–410
 - gate, 498–500
 - sheet, 14
 - source series, 296–298
- Resistivity, 13

Reverse bias, 26, 29–31
 Reverse-breakdown voltage, 30
 Reverse saturation, 146, 155–156
 Reverse short-channel effect (RSCE), 269
 RF modeling
 cutoff (transition) frequency, 500–502
 gate resistance, 498–500
 maximum frequency of oscillation
 (unity power gain frequency),
 502–504
 model topologies, 495–498
 Root mean square (rms) value, 411
 RSCE (reverse short-channel effect), 269

S

Salicide (self-aligned silicide) process,
 298
 Saturation
 forward, 154–155
 reverse, 146, 155–156
 Saturation current, 30
 Saturation region, 40, 41, 44–46, 129,
 129n, 146, 154–156, 427–428
 Scaling, 290–296
 Scattering
 Coulomb, 181, 182
 introduced, 9–10
 phonon, 181
 Self-aligned silicide (salicide) process,
 298
 Semiconductors
 degenerate, 4
 energy bands in, 557–561
 extrinsic, 3–7, 558
 intrinsic, 2–3, 557–558
 introduced, 1–8
 n type, 4
 p type, 5
 Shallow-trench isolation (STI) CMOS
 process, 37, 38, 270, 271, 272,
 275–277
 Sheet resistance, 14
 Short-channel devices, 257–269
 Shot noise, 421
 Silicon on insulator (SOI) CMOS process,
 37, 38
 Simulation; *see* Circuit simulation,
 models for
 Small-dimension effects
 barrier lowering, two-dimensional
 charge sharing, and threshold
 voltage
 introduced, 257
 narrow-channel devices, 270–277
 short-channel devices, 257–269
 summary and comments, 277
 carrier velocity saturation, 280–286
 channel-length modulation (CLM),
 249–257
 effect of source and drain series
 resistances, 296–298
 effects due to thin oxides and high
 doping, 298–301
 hot carrier effects, 286–290
 intrinsic capacitances and, 402–405
 introduced, 248–249
 noise and, 424–425
 punchthrough, 277–279
 scaling, 290–296
 Small-signal capacitances; *see*
 Capacitances
 Small-signal modeling for high
 frequencies
 high-frequency noise, 492–495
 introduced, 440
 non-quasi-static models
 introduced, 467–468
 model comparisons, 491–492
 other approximations and higher-
 order models, 490–491
 strong-inversion model; *see* Non-
 quasi-static strong-inversion model
 quasi-static model, complete
 description of capacitance effects,
 440–445

- equivalent circuit topologies, 445–451
- evaluation of capacitances, 451–458, 459–460
- frequency region of validity, 458, 460
- RF modeling
 - cutoff (transition) frequency, 500–502
 - gate resistance, 498–500
 - maximum frequency of oscillation (unity power gain frequency), 502–504
 - model topologies, 495–498
 - y-parameter models, 460–467, 477–489, 606–608
- Small-signal modeling for low and medium frequencies
 - extrinsic part, 405–410
 - general models, 427–428
 - introduced, 359–360
 - low-frequency, intrinsic part
 - channel path, 361–365
 - drain-to-substrate path, 365–367
 - general models, 381–384, 385
 - moderate inversion, 380–381
 - strong inversion, 367–368
 - two-path view, 360–361
 - weak inversion, 378–380
 - medium-frequency, intrinsic part
 - intrinsic capacitances; *see* Intrinsic capacitances
 - introduced, 384–385
- noise
 - equivalent-circuit model, 425–426
 - equivalent input noise resistance, 420
 - equivalent noise voltage, 420
 - flicker noise, 413, 422–424, 539
 - introduced, 410–414
 - random telegraph noise, 424–425
 - shot noise, 421
 - small-dimension effects, 424–425
 - thermal (Johnson, Nyquist) noise, 412, 414, 539
 - white noise, 412, 413, 414–422
- SOI (silicon on insulator) CMOS process, 37, 38
- Source, 35, 36
- Source-drain conductance, 363, 370–375
- Source-drain proximity capacitance, 409
- Source-referenced versus body-referenced modeling, 179–181
- Source series resistance, 296–298
- Step (abrupt) junctions, 26–27
- STI (shallow-trench isolation) CMOS process, 37, 38, 270, 271, 272, 275–277
- Strong inversion
 - in depletion *n*MOS transistors, approximate strong-inversion model for *n*-implant-*p*-substrate devices, 238
 - in enhancement *n*MOS transistors implanted
 - drain-to-source current model, 219–222
 - simplified model, 222–224
 - in four-terminal MOS transistors, principles
 - charges for, 603–605
 - complete symmetric strong-inversion model, 150–156
 - introduced, 40, 45, 46, 150
 - model origin summary, 169–170
 - simplified, source-referenced, strong-inversion model, 158–169
 - simplified symmetric strong-inversion model, 156–158
 - intrinsic capacitances and, 390–393
 - in low-frequency model for the intrinsic part, 367–368
 - in *n*MOS transistors, 40, 45, 46
 - quasi-static operation and, 325–331, 603–605
 - in three-terminal MOS structures, 106–109, 121
 - in two-terminal MOS structures, 71–74, 87

Strong-inversion model, non-quasi-static;
 see Non-quasi-static strong-
 inversion model
 Substrate (body), 34, 35
 Substrate (body) effect; *see* Body
 (substrate) effect
 Substrate (body) terminal, 51, 52
 Substrate transconductance, 363, 368–370
 Surface, 34
 Surface-channel devices, 226
 Surface-channel enhancement-mode
 pMOS, 240–241
 Surface mobility, 132, 181–182, 237
 Surface potentials, 56–58
 Surface punchthrough, 278–279

T

Table models, 514–515
 Temperature effects, 189–191
 Thermal (Johnson, Nyquist) noise, 412,
 414, 539
 Thermal voltage, 6–7
 Three-terminal MOS structures
 body (substrate) effect, 101–103
 energy bands in, 587–590
 general analysis, 591–593
 introduced, 91
 inversion
 approximate limits of regions,
 103–106
 moderate inversion, 113, 122, 586
 strong inversion, 106–109, 121
 weak inversion, 109–113, 121
 V_{CB} control point of view
 fundamentals, 113–117, 118
 pinchoff voltage, 117–122
 Threshold voltage, 72–73, 108, 159,
 160–161, 169n, 196–197; *see also*
 Effective threshold voltage
 Transadmittance, high-frequency,
 538–539
 Transconductance; *see also* Conductance
 apparent transconductance overshoot,
 378
 gate, 361, 367–368
 substrate, 363, 368–370
 transconductance-to-current ratio,
 534–535, 536
 Transconductance-to-current ratio,
 534–535, 536
 Transit time, 8–9, 13, 17–18
 quasi-static operation and, 339–340, 341
 Transition (cutoff) frequency, 500–502
 Transport current, 313–314
 Triode region, 155n
 Tunneling, 289, 300–301
 Two-terminal MOS structures
 charge density, electric field, and
 potential in, 578–579
 effect of gate-substrate voltage on
 surface condition
 accumulation, 59, 60
 depletion and inversion, 59–62
 flat-band condition, 59
 general analysis, 62–64, 67
 introduced, 58
 energy bands in, 574–577
 flat-band voltage, 51–56, 59
 general analysis, 580–583
 introduced, 50–51
 inversion
 general relations and regions, 64–71,
 582–583
 moderate inversion, 78–79, 87,
 584–585
 strong inversion, 71–74, 87
 summary of properties, 87
 weak inversion, 74–78, 87
 potential balance and charge balance,
 56–58, 67
 small-signal capacitance, 79–86

U

Unilateral power gain, 502–503
 Unity gain power frequency (maximum frequency of oscillation), 502–504

V

Vacuum energy level, 561, 562
 Valence electrons, 2
 Velocity overshoot, 280n
 Velocity saturation effects, 280–286

W

Weak avalanche, 286
 Weak inversion
 in enhancement *n*MOS transistors, 225

 in four-terminal MOS structures, 170–175
 intrinsic capacitances and, 399–400
 in low-frequency model for the intrinsic part, 378–380
 quasi-static operation and, 332–333
 in three-terminal MOS structures, 109–113, 121
 in two-terminal MOS structures, 74–78, 87
 White noise, 412, 413, 414–422
 Work functions, 22n, 53n, 561–562

Y

y-parameter models, 460–467, 477–489, 606–608